

# Branching Process Models of Cancer

Richard Durrett, Duke U.

December 7, 2014

## Abstract

In this chapter, we will use multitype branching processes with mutation to model cancer. With cancer progression, resistance to therapy, and metastasis in mind, we will investigate  $\tau_k$ , the time of the first type  $k$  mutation, and  $\sigma_k$ , the time of the first type  $k$  mutation that founds a family line that does not die out, as well as the growth of the number of type  $k$  cells. The last three sections apply these results to metastasis, ovarian cancer, and tumor heterogeneity. Even though martingales and stable laws are mentioned, these notes should be accessible to a student who is familiar with Poisson processes and continuous time Markov chains.

## 1 Multistage theory of cancer

Throughout these notes, we model cancer as an exponentially growing cell population in which type  $i$  cells are those that have accumulated  $i \geq 0$  mutations compared to the type 0 cells. To motivate the study of these models we begin with a brief and incomplete history lesson and a description of some situations in which our model may be applied.

The idea that carcinogenesis is a multistage process goes back to the 1950s. Fisher and Holloman [52] pointed out that when the logarithm of cancer mortality from stomach cancer in women was plotted versus log of age, the result was line with slope 6. Nordling [65] and Armitage and Doll [43] suggested that the observed relationship would be explained if a cancer cell was the end result of seven mutations. There was no model. The conclusion was based on the fact that if  $X_i$  are independent and exponential with rate  $u_i$  then

$$P(X_1 + \cdots + X_k \leq t) \sim \left( \prod_{i=1}^k u_i \right) \frac{t^{k-1}}{(k-1)!},$$

the restriction to small  $t$  being due to the fact that most cancers affect only a small fraction of the population. For more on early work, see the survey by Armitage [42].

In 1971 Knudson [55] analyzed 48 cases of retinoblastoma, a cancer that develops in the retinas of children as the eye grows its full size in the first five years of life. 25 children had tumors in only one eyes, while 23 had tumors in both eyes and generally more serious symptoms. From the age of onset of cancer for patients in the two groups, he inferred that there was a gene, later identified and named *RB1*, so that when both copies were knocked out cancer was initiated, and that individuals in the group with bilateral tumors had a germline mutation so that one copy was knocked out in all of their cells. This was the first example of a *tumor suppressor gene*, which leads to trouble when both copies have been inactivated. For more see Knudson's 2001 survey [56].

In the late 1980s, it became possible to identify the molecular events that underlie the initiation and progression of human tumors. The abundant data for colorectal tumors made it an excellent system to study the genetic alterations involved. Fearon and Vogelstein [49] argued that mutations in four or five genes were required for the formation of a malignant tumor, and that colorectal tumors appear to arise from the mutational inactivation of tumor suppressor genes coupled with a mutational activation of *oncogenes*; one hit turns them on. In population genetics terms the later would be called advantageous mutations because they increase the growth rate of the cells in which they occur.

Over a decade later, Luebeck and Moolgavkar [60] synthesized the existing research on colon cancer to produce a four stage model. The first two steps were the knockout of the adenomatous polyposis coli (*APC*) gene, followed by activation of the oncogene *KRAS* and the inactivation of *TP53*, which has been called “the guardian of the genome” because of its role in conserving stability by preventing genome mutation. For more on the genetic basis of colon cancer, see Fearon's recent survey [48].

Compared to colon cancer, chronic myeloid leukemia (*CML*) is a very simple disease. 95% of patients have a *BCR-ABL* gene fusion caused by a reciprocal translocation between chromosome 9 and chromosome 22. Because of this simplicity, the disease can be treated with a tyrosine kinase inhibitor, e.g., imatinib, that blocks the action of *BCR-ABL*. Unfortunately, mutations in the *ABL* binding domain can lead to resistance to therapy. In Sections 6 and 12, we will use a two type branching process model to investigate the questions: What is the probability that resistant cells are present at diagnosis? If so, how many are there?

Metastasis, the spread of cancer to distant organs, is the most common cause of death for cancer patients. It is a very complex process: cells must enter the blood stream (intravasation), survive the trip through the circulatory system, leave the blood stream at its destination (extravasation), and survive in an alien environment, e.g., cells from the breast tissue living in bone. See [76] and [50] for more detailed descriptions. In Section 16, we will analyze a simple three type model for metastasis that ignores the most of the details of the process. In Section 17 we will consider the special case of ovarian cancer.

## 2 Mathematical Overview

Our model is a multitype branching process with mutation in which  $Z_i(t)$  be the number of type  $i$  cells at time  $t$ . Type  $i$  cells give birth at rate  $a_i$  and die at rate  $b_i$ . Here, we always assume that the growth rate  $\lambda_i = a_i - b_i > 0$ , even though this does not hold in all applications, see e.g., [16, 17]. 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  $i + 1$ . Some researchers prefer to have births at rate  $\alpha_i$  and mutations at birth with probability  $\mu_i$ , but this is equivalent to  $a_i = \alpha_i(1 - \mu_i)$  and  $u_i = \alpha_i\mu_i$ .

For our model it is natural to investigate:

- $\tau_k$ , the time of the first type  $k$  mutation, and  $\sigma_k$ , the time of the first type  $k$  mutation that founds a family line that does not die out. The limiting distributions of these quantities as  $s, u_i \rightarrow 0$  are studied in Section 5 for the case  $k = 1$  and in general in Section 15.
- The limiting behavior of  $e^{-\lambda_1 t} Z_1(t)$  as  $t \rightarrow \infty$  is studied in Section 9, where one sided-stable laws with index  $\alpha < 1$ . Once the result is proved for  $k = 1$ , it extends in a straightforward way to  $k > 1$ , see Section 13. The method of proof of these results allows us to obtain insights into tumor heterogeneity in Section 18.
- Let  $T_M$  the time for the type 0 population to reach size  $M$ , which we think of the time at which cancer is detected. Motivated by the questions about resistance to imatinib, we will investigate  $P(Z_1(T_M) > 0)$  in Section 6 and the size of  $Z_1(T_M)$  in Section 12. Thinking about metastasis, we will study  $P(Z_2(T_M) > 0)$  in Section 16.

There are more results in this survey, but these are the main themes. Along the way we will review, and in some cases improve, results that have been derived using nonrigorous arguments.

Some of our formulas are complicated, so it useful to look at concrete examples. To do this, we need to have reasonable parameter values. We will typically take the death rates  $b_i = 1$ . This corresponds to measuring time in units of cell divisions. The exponential growth rate  $\lambda_i$  is analogous to the selective advantage of type  $i$  cells. In a study of glioblastoma and colorectal cancer, Bozic et al [4] concluded that the average selective advantage of somatic mutations was surprisingly small, 0.004. Here, we will often use slightly larger values,  $\lambda_1 = 0.02$ , and  $\lambda_2 = 0.04$ .

To identify the order of magnitude of the mutation rates  $u_i$ , we note that the point mutation rate has been estimated, see [58], to be  $5 \times 10^{-10}$  per nucleotide per cell division. To compute the  $u_i$  this number needs to be multiplied by the number of nucleotides that when mutated lead to cancer. In some cases there are a small number of nonsynonymous mutations (that change the amino acid in the

corresponding protein) that achieve the desired effect, while in other cases there may be hundreds of possible mutations that knock out the gene and there may be a number of genes within a genetic pathway that can be hit. Thus mutation rates can range from  $10^{-9}$  to  $10^{-5}$ , or can be larger after the mechanisms that govern DNA replication are damaged.

### 3 Branching process results

We begin by studying the number of type 0 cells,  $Z_0(t)$ , which is a branching process in which each cell gives birth at rate  $a_0$  and dies at rate  $b_0$ . In terms of the theory of continuous time Markov chains, the matrix  $q(i, j)$  that gives the rate of jumps from  $i$  to  $j$  has

$$q(i, i + 1) = a_0 i, \quad q(i, i - 1) = b_0 i,$$

and  $q(i, j) = 0$  otherwise. Since each initial individual gives rise to an independent copy of the branching process, we will suppose throughout this section that  $Z_0(0) = 1$ . Since each individual gives birth at rate  $a_0$  and dies as rate  $b_0$

$$\frac{d}{dt} E Z_0(t) = \lambda_0 E Z_0(t),$$

where  $\lambda_0 = a_0 - b_0$ . Since  $E Z_0(0) = 1$ ,

$$E Z_0(t) = e^{\lambda_0 t}. \tag{1}$$

Our next step is to compute the extinction probability,

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

By considering what happened on the first jump

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

In words, since jumps  $1 \rightarrow 0$  occur at rate  $b_0$  and from  $1 \rightarrow 2$  at rate  $a_0$ , the first event is a death with probability  $b_0/(a_0 + b_0)$  in which case the probability of extinction is 1. The first event is a birth with probability  $a_0/(a_0 + b_0)$  in which case the probability of extinction is  $\rho^2$ , since for this to happen both of the lineages have to die out and they are independent.

Rearranging (2) 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

$$\rho = \begin{cases} b_0/a_0 & \text{if } a_0 > b_0, \\ 1 & \text{if } a_0 \leq b_0. \end{cases} \tag{3}$$

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

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

*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  $\approx a_0h$ , of a death is  $\approx b_0h$ . 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.  $\square$

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

$$F(x, t) = \frac{b_0(x - 1) - e^{-\lambda_0 t}(a_0x - b_0)}{a_0(x - 1) - e^{-\lambda_0 t}(a_0x - b_0)}. \quad (4)$$

**Remark.** Here and in what follows, if the reader finds the details of the calculations too unpleasant, feel free to skip ahead to the end of the proof.

*Proof.* Later we will need a generalization so we will find the solution of

$$\frac{dg}{dt} = (c - g)(b - ag) \quad g(0) = x.$$

The equation in Lemma 1 corresponds to  $c = 1$ ,  $a = a_0$ , and  $b = b_0$ . Rearranging we have

$$dt = \frac{dg}{(c - g)(b - ag)} = \frac{1}{b - ac} \left( \frac{dg}{c - g} - a \frac{dg}{b - ag} \right).$$

Integrating we have for some constant  $D$

$$t + D = -\frac{1}{b - ac} \log(c - g) + \frac{1}{b - ac} \log(b - ag) = \frac{1}{b - ac} \log \left( \frac{b - ag}{c - g} \right), \quad (5)$$

so  $b - ag = e^{(b-ac)t} e^{D(b-ac)}(c - g)$ , and solving for  $g$  gives

$$g(t) = \frac{b - ce^{D(b-ac)}e^{(b-ac)t}}{a - e^{D(b-ac)}e^{(b-ac)t}}. \quad (6)$$

Taking  $t = 0$  in (6), we have

$$e^{D(b-ac)} = \frac{ax - b}{x - c}.$$

Using this in (6), setting  $c = 1$ , adding the subscript 0's, and rearranging gives (4).  $\square$

**Remark.** Here and throughout the paper we use  $\log$  for the natural (base  $e$ ) logarithm.

It is remarkable that one can invert the generating function to find the underlying distribution. These formulas come from Section 8.6 in Bailey (1964). If we let

$$\alpha = \frac{b_0 e^{\lambda_0 t} - b_0}{a_0 e^{\lambda_0 t} - b_0} \quad \text{and} \quad \beta = \frac{a_0 e^{\lambda_0 t} - a_0}{a_0 e^{\lambda_0 t} - b_0}, \quad (7)$$

then the underlying distribution is a generalized geometric

$$p_0 = \alpha \quad p_n = (1 - \alpha)(1 - \beta)\beta^{n-1} \quad \text{for } n \geq 1. \quad (8)$$

To check this claim, note that taking  $x = 0$  in (4) confirms the size of the atom at 0 and suggests that we write

$$F(x, t) = \frac{(b_0 e^{\lambda_0 t} - b_0) - x(b_0 e^{\lambda_0 t} - a_0)}{(a_0 e^{\lambda_0 t} - b_0) - x(a_0 e^{\lambda_0 t} - a_0)}. \quad (9)$$

For comparison we note that the generating function of (8) is

$$\begin{aligned} \alpha + (1 - \alpha)(1 - \beta) \sum_{n=1}^{\infty} \beta^{n-1} x^n &= \alpha + (1 - \alpha)(1 - \beta) \frac{x}{1 - \beta x} \\ &= \frac{\alpha + (1 - \alpha - \beta)x}{1 - \beta x}. \end{aligned}$$

Dividing the numerator and the denominator in (9) by  $a_0 e^{\lambda_0 t} - b_0$  gives

$$F(x, t) = \frac{\alpha - x(b_0 e^{\lambda_0 t} - a_0)/(a_0 e^{\lambda_0 t} - b_0)}{1 - \beta x},$$

and it remains to check that using (7) that

$$1 - \alpha - \beta = \frac{(a_0 - b_0)e^{\lambda_0 t} - a_0 e^{\lambda_0 t} + a_0}{a_0 e^{\lambda_0 t} - b_0} = -\frac{b_0 e^{\lambda_0 t} - a_0}{a_0 e^{\lambda_0 t} - b_0}.$$

**Yule process.** If  $b_0 = 0$  (no death), and hence  $\lambda_0 = a_0$ , then the generating function in (4) simplifies to

$$F(x, t) = \frac{x e^{-\lambda_0 t}}{1 - x + x e^{-\lambda_0 t}}. \quad (10)$$

$\alpha = 0$  in (8), so we have a geometric distribution with

$$\beta = 1 - e^{-\lambda_0 t}.$$

Since the mean of the geometric is  $1/(1 - \beta)$ , this is consistent with the fact that  $EZ_0(t) = e^{\lambda_0 t}$

Returning to the general case and subtracting (4) from 1, we have

$$1 - F(x, t) = \frac{\lambda_0(x - 1)}{a_0(x - 1) - e^{-\lambda_0 t}(a_0 x - b_0)}. \quad (11)$$

Setting  $x = 0$  in (11), we have

$$\begin{aligned} P(Z_0(t) = 0) &= \frac{b_0 - b_0 e^{-\lambda_0 t}}{a_0 - b_0 e^{-\lambda_0 t}} \quad \text{and} \\ P(Z_0(t) > 0) &= 1 - F(0, t) = \frac{\lambda_0}{a_0 - b_0 e^{-\lambda_0 t}}. \end{aligned} \quad (12)$$

Note that this converges exponentially fast to the probability of nonextinction,  $\lambda_0/a_0$ .

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

$$W_0 =_d \frac{b_0}{a_0} \delta_0 + \frac{\lambda_0}{a_0} \text{exponential}(\lambda_0/a_0) \quad (13)$$

where  $\delta_0$  is a point mass at 0, and the exponential( $r$ ) distribution has density  $re^{-rt}$  and mean  $1/r$ . That is,

$$P(W_0 = 0) = \frac{b_0}{a_0} \quad \text{and} \quad P(W_0 > x | W_0 > 0) = \exp(-x\lambda_0/a_0).$$

*Proof.* From (8), the atom at 0

$$\alpha(t) = \frac{b_0 - b_0 e^{-\lambda_0 t}}{a_0 - b_0 e^{-\lambda_0 t}} \rightarrow \frac{b_0}{a_0}.$$

The geometric distribution  $(1 - \beta)\beta^{n-1}$  has mean

$$\frac{1}{1 - \beta} = \frac{a_0 e^{\lambda_0 t} - b_0}{a_0 - b_0} \sim \frac{a_0 e^{\lambda_0 t}}{\lambda_0},$$

where  $x_t \sim y_t$  means  $x_t/y_t \rightarrow 1$ . Since a geometric distribution rescaled by its mean converges to a mean 1 exponential the desired result follows.

This gives the limiting behavior of the distribution of  $e^{-\lambda_0 t} Z_0(t)$ . To show that the sequence of numbers  $e^{-\lambda_0 t} Z_0(t) \rightarrow W_0$ , we note that  $E Z_0(t) = e^{\lambda_0 t}$  and individuals give birth independently, so  $e^{-\lambda_0 t} Z_0(t)$  is a nonnegative martingale and hence converges with probability one to a limit  $W_0$ .  $\square$

If we let  $\Omega_0^0 = \{Z_0(t) = 0 \text{ for some } t \geq 0\}$  then (3) implies

$$P(\Omega_0^0) = b_0/a_0. \quad (14)$$

Since  $W_0 = 0$  on  $\Omega_0^0$ , (13) 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

$$(e^{-\lambda_0 t} Z_0(t) | \Omega_\infty^0) \rightarrow V_0 = \text{exponential}(\lambda_0/a_0). \quad (15)$$

Later we need the formula for the Laplace transform of  $V_0$ :

$$\begin{aligned} Ee^{-\theta V_0} &= \int_0^\infty e^{-\theta x} \frac{\lambda_0}{a_0} e^{-(\lambda_0/a_0)x} dx \\ &= \frac{\lambda_0/a_0}{(\lambda_0/a_0) + \theta} = (1 + (a_0/\lambda_0)\theta)^{-1}. \end{aligned} \quad (16)$$

### 3.1 Conditioned branching processes

For almost all of these notes, we will consider processes in continuous time. The one exception is that to develop the theory in this section, we begin with discrete time. In that setting the branching process is described by giving the offspring distribution  $p_k$ , i.e., the probability an individual has  $k$  children. Given a family of independent random variables  $\xi_i^n$  with  $P(\xi_i^n = k) = p_k$  we can define the branching process by specifying  $Z_0$  and then inductively defining

$$Z_{n+1} = \begin{cases} 0 & \text{if } Z_n = 0, \\ \sum_{i=1}^{Z_n} \xi_i^n & \text{if } Z_n > 0. \end{cases}$$

Let  $\mu = \sum_k k p_k$  be the mean of the offspring distribution. It is known that if  $\mu < 1$  or  $\mu = 1$  and  $p_1 < 1$  the process always dies out, while if  $\mu > 1$  then the probability the process dies out starting from  $Z_0 = 1$  is the solution  $\in [0, 1)$  of  $\phi(\rho) = \rho$ , where  $\phi(x) = \sum_k p_k x^k$  is the generating function of the offspring distribution.

In some situations, it is desirable to decompose a branching process with  $\mu > 1$  into the backbone (the individuals who have an infinite line of descent) and the side trees (the individuals who start a family that dies out). A fact, due to Harris [23], which makes this useful, is that if we condition on the branching process not dying out, and only look at the individuals that have an infinite line of descent, then we have a branching process with offspring distribution:

$$\bar{p}_k = (1 - \rho)^{-1} \sum_{m \geq k} p_m \binom{m}{k} (1 - \rho)^k \rho^{m-k}. \quad (17)$$

In words, if we want  $k$  children whose family lines do not die out then we need to have  $m \geq k$  children and exactly  $m - k$  of these must start lineages that die out. We divide by  $1 - \rho$  because that is the probability that the individual under consideration starts a family line that does not die out.



To compute the generating function of  $\bar{p}_k$ , we interchange the order of summation, which is legitimate because all the terms are  $\geq 0$ , and then use the binomial formula to evaluate the inner sum.

$$\begin{aligned} \sum_{k=1}^{\infty} \sum_{m \geq k} p_m \binom{m}{k} (1-\rho)^k \rho^{m-k} z^k &= \sum_{m=1}^{\infty} p_m \sum_{k=1}^m \binom{m}{k} (1-\rho)^k z^k \rho^{m-k} \\ &= \sum_{m=1}^{\infty} p_m [(1-\rho)z + \rho]^m - \rho^m, \end{aligned}$$

so the generating function is

$$\bar{\phi}(z) = \sum_{k=1}^{\infty} \bar{p}_k z^k = \frac{\phi((1-\rho)z + \rho) - \phi(\rho)}{1-\rho}.$$

Geometrically, we have taken the part of the generating function over  $[\rho, 1]$ , which has range  $[\rho, 1]$ , and linearly rescaled the  $x$  and  $y$  axes to make it map  $[0, 1]$  onto  $[0, 1]$ . See Figure 1 for a picture.

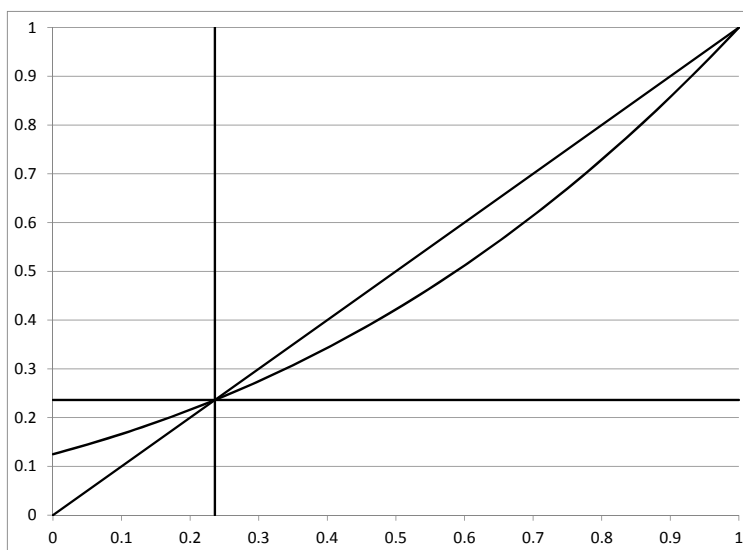


Figure 1: Graph of the generating function  $\phi(z) = (1+z)^3/8$  for Binomial(3,1/2), showing  $\bar{\phi}$  in the upper right square and  $\hat{\phi}$  in the lower left. Fixed point is at  $-2+\sqrt{5} = 0.23606$ .

If we condition our branching process to die out, we get a branching process with offspring distribution:

$$\hat{p}_k = p_k \rho^k / \rho. \quad (18)$$

In words, if there are  $k$  children then all of their family lines must die out, and these events are independent with probability  $\rho$ . We divide by  $\rho$  because that if the

probability that the individual under consideration starts a family line that dies out. The generating function is

$$\hat{\phi}(z) = \sum_{k=1}^{\infty} p_k \rho^k z^k / \rho = \phi(\rho z) / \rho.$$

In this case, we have taken the part of the generating function over  $[0, \rho]$ , which has range  $[0, \rho]$ , and linearly rescaled the  $x$  and  $y$  axes to make it map  $[0, 1]$  onto  $[0, 1]$ . See Figure 1 for a picture. The mean of  $\hat{p}_k$  is  $\hat{\mu} = \sum_{k=1}^{\infty} k p_k \rho^{k-1} = \phi'(\rho)$ , so the mean total progeny of a process conditioned to die out is

$$\sum_{n=0}^{\infty} \phi'(\rho)^n = 1 / (1 - \phi'(\rho)). \quad (19)$$

To do these conditionings in continuous time, we follow O'Connell [33] and take a limit of discrete time approximations. Let the offspring distribution be

$$p_0 = b_0 \delta \quad p_2 = a_0 \delta \quad p_1 = 1 - (a_0 + b_0) \delta$$

where  $\delta$  is small enough so that  $p_1 > 0$ . The extinction probability solves

$$\rho = b_0 \delta + [1 - (a_0 + b_0) \delta] \rho + a_0 \delta \rho^2,$$

or  $b_0 - (a_0 + b_0) \rho + a_0 \rho^2 = 0$  which again gives  $\rho = b_0 / a_0$  if  $a_0 > b_0$ . If we condition the branching process to die out then by (18)

$$\hat{p}_0 = p_0 / \rho = a_0 \delta \quad \hat{p}_2 = p_2 \rho b_0 \delta \quad \hat{p}_1 = p_1 = 1 - (a_0 + b_0) \delta$$

so in the limit as  $\delta \rightarrow 0$  we have births at rate  $b_0$  and deaths at rate  $a_0$ . If we condition on survival and recall  $\rho = a_0 / b_0$ , then we get

$$\bar{p}_2 = p_2(1 - \rho) = b_0 \delta(1 - \rho) = \lambda_0 \delta \quad \bar{p}_1 = 1 - \bar{p}_2 = 1 - \lambda_0 \delta$$

so in the limit as  $\delta \rightarrow 0$  we end up with a Yule process with births at rate  $\lambda_0$ .

## 4 Time for $Z_0$ 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  $M$ . For example, in studies of chronic myeloid leukemia it has been common to take  $M = 10^5$  cancerous cells.

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 (15), conditional on nonextinction,  $e^{-\lambda_0 t} Z_0(t) \rightarrow V_0$ , which is exponential with rate  $\lambda_0/a_0$ , or informally  $Z_0(t) \approx e^{\lambda_0 t} V_0$ . From this we see that

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

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

$$f_{T_M}(t) = \exp(-(\lambda_0/a_0) M e^{-\lambda_0 t}) \cdot \frac{\lambda_0^2 M}{a_0} e^{-\lambda_0 t} \quad (20)$$

Clearly the actual waiting  $T_M \geq 0$ , however in our formula  $P(T_M \leq 0) = \exp(-\lambda_0 M/a_0)$ . As we will see in the concrete example below, in applications this will be very small. Because of this, it is natural to view the density in (20) as defined on  $(-\infty, \infty)$ . To compute the mean we have to compute

$$ET_M = \frac{\lambda_0^2 M}{a_0} \int_{-\infty}^{\infty} t e^{-\lambda_0 t} \exp(-(\lambda_0/a_0) M e^{-\lambda_0 t}) dt$$

If we let  $z = (\lambda_0/a_0) M e^{-\lambda_0 t}$  then  $e^{-\lambda_0 t} = a_0 z / \lambda_0 M$ ,  $t = -(1/\lambda_0) \log(a_0 z / \lambda_0 M)$  and  $dt = -dz / z \lambda_0$  so the integral becomes

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

Since  $\int_0^{\infty} e^{-z} dz = 1$  it follows that

$$ET_M = \frac{1}{\lambda_0} \log\left(\frac{M \lambda_0}{a_0}\right) - \frac{1}{\lambda_0} \int_0^{\infty} e^{-z} \log z dz \quad (21)$$

The first term is the value of  $T_M$  if we replace  $V_0$  by its mean  $a_0/\lambda_0$  and solve

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

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

$$\gamma = 0.5772156649.$$

**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 (21) is

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

and the second is  $\gamma/\lambda_0 = 28.86$  a small correction.  $b = 1$  so the units are the number of cell divisions, i.e., for cells that divide on the average every four days this is 1516 days or 4.15 years.

## 5 Time until the first type 1

Let  $\tau_1$  be the time of occurrence of the first type 1. If we think of 0's as an exponentially growing population of tumor cells then type 1's cells might be have more aggressive growth or be resistant to therapy. Since 1's are produced at rate  $u_1 Z_0(s)$  at time  $s$ ,

$$P(\tau_1 > t | Z_0(s), s \leq t, \Omega_\infty^0) = \exp\left(-u_1 \int_0^t Z_0(s) ds\right). \quad (22)$$

$\tau_1$  will typically occur when  $\int_0^t Z_0(s) ds$  is of order  $1/u_1$ . A typical value for the mutation rate is  $u_1 = 10^{-5}$  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,

$$\int_0^t e^{\lambda_0 s} V_0 ds = V_0 \cdot \frac{e^{\lambda_0 t} - 1}{\lambda_0}.$$

Dropping the second term and taking the expected value by using (16) with  $\theta = u_1 e^{\lambda_0 t} / \lambda_0$ , we conclude that

$$P(\tau_1 > t | \Omega_\infty^0) \approx E \exp(-u_1 V_0 e^{\lambda_0 t} / \lambda_0) = (1 + (a_0 / \lambda_0^2) u_1 e^{\lambda_0 t})^{-1}. \quad (23)$$

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

$$t_{1/2}^1 \approx \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2}{a_0 u_1}\right). \quad (24)$$

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

$$P(\tau_1 > t | V_0) \approx \exp(-u_1 V_0 e^{\lambda_0 t} / \lambda_0). \quad (25)$$

If we replace  $V_0$  by its mean  $EV_0 = a_0 / \lambda_0$  the tail of the limit distribution of  $\tau_1$  is equal to  $1/e$  at

$$\bar{t}_{1/e}^1 = \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2}{a_0 u_1}\right). \quad (26)$$

A second quantity of interest is  $\sigma_1$ , 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 (23) and (25) to get

$$P(\sigma_1 > t | \Omega_\infty^0) \approx (1 + (a_0 / \lambda_0^2) (u_1 \lambda_1 / a_1) e^{\lambda_0 t})^{-1}, \quad (27)$$

$$P(\sigma_1 > t | V_0) \approx \exp(-V_0 (u_1 \lambda_1 / a_1) e^{\lambda_0 t} / \lambda_0). \quad (28)$$

Replacing  $t$  by  $s$  to define the quantities for  $\sigma_1$  corresponding to (24) and (26)

$$s_{1/2}^1 = \bar{s}_{1/e}^1 = \frac{1}{\lambda_0} \log\left(\frac{\lambda_0^2 a_1}{a_0 u_1 \lambda_1}\right). \quad (29)$$

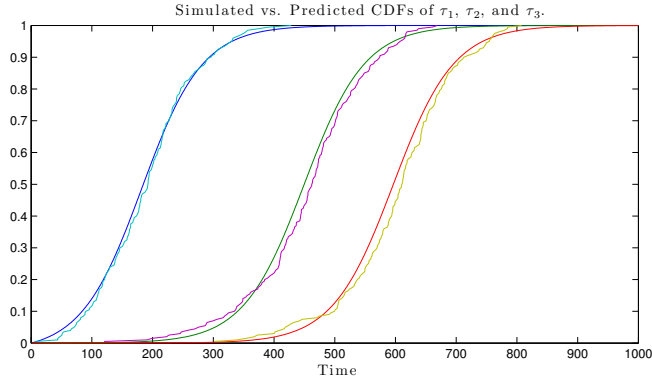


Figure 2: Results of 200 runs of the system with  $a_0 = 1.02$ ,  $a_1 = 1.04$ ,  $a_2 = 1.06$ ,  $b_i = 1.0$ , and  $u = 10^{-5}$ . Smooth curves are the limit results for  $\tau_i$ ,  $i = 1, 2, 3$ .

**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

$$t_{1/2}^1 = \bar{t}_{1/e}^1 = 50 \log \left( \frac{4 \times 10^{-4}}{1.02 \times 10^{-5}} \right) = 183.45,$$

$$s_{1/2}^1 = \bar{s}_{1/e}^1 = 50 \log \left( \frac{4.16 \times 10^{-4}}{4.08 \times 10^{-7}} \right) = 50 \log(1019.6) = 346.36.$$

Again since  $b_i = 1$ , the time units are cell-divisions. If cells divide every four days, this translates into 733.8 days (about two years) and 1385.44 days or 3.8 years. If instead  $u_1 = 10^{-6}$  this adds  $(1/\lambda_0) \log(10) = 115.13$  or 460.5 days to the two waiting times.

**Limit Theorems.** Our next goal is to find the limiting behavior of  $\tau_1$ . Since the median is where the distribution function crosses  $1/2$ , (22) implies

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

and it follows that

$$P(\tau_1 > t_{1/2}^1 + x/\lambda_0 | \Omega_\infty^0) \rightarrow (1 + e^x)^{-1}. \quad (30)$$

For a comparison with simulation see Figure 2.

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

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

The results for  $\sigma_1$  come from changing the value of  $u_1 \rightarrow u_1 \lambda_2 / a_2$ .

$$\begin{aligned} P(\sigma_1 > s_{1/2}^1 + x/\lambda_0 | \Omega_\infty^0) &= P(\tau_1 > t_{1/2}^1 + x/\lambda_0 | \Omega_\infty^0) \rightarrow (1 + e^x)^{-1}, \\ P(\sigma_1 > \bar{s}_{1/e}^1 + x/\lambda_0 | V_0) &= P(\tau_1 > \bar{t}_{1/e}^1 + x/\lambda_0 | V_0) \rightarrow \exp(-e^x). \end{aligned} \quad (31)$$

## 6 Mutation before detection?

Cancer therapy often fails because acquired resistance enables cancer cells to grow despite the continuous administration of therapy. In some cases, a single genetic alteration is sufficient to cause resistance to cancer treatment. Chronic myeloid leukemia (CML) is caused by a reciprocal translocation between chromosomes 9 and 22 which creates the BCR-ABL oncoprotein. Treatment of CML with imatinib can fail due to a single point mutation in the tyrosine kinase domain of ABL. To date more than 90 point mutations that cause resistance to treatment have been observed. The possibility of drug-resistant cells at the beginning of therapy is of clinical importance since the likelihood and extent of resistance determines treatment choices and patient prognosis.

For this reason Iwasa, Nowak, and Michor [24], 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, we assume that the disease can be detected when the number of cancer cells reaches size  $M$ . Assuming that the number of resistant (type 1) cells will be a small fraction of  $M$  we will define this to be  $T_M = \min\{t : Z_0(t) = M\}$ . Using the calculation in (22), and noting that on the nonextinction event  $\Omega_\infty^0$ , we have  $Z_0(t) \sim V_0 e^{\lambda_0 t}$  which implies  $Z_0(T_M - s) \approx M e^{-\lambda_0 s}$ , we find

$$\begin{aligned} P(\tau_1 > T_M | Z_0(s), s \leq T_M, \Omega_\infty^0) &= \exp\left(-u_1 \int_0^{T_M} Z_0(t) dt\right) \\ &\approx \exp\left(-M u_1 \int_0^\infty e^{-\lambda_0 s} ds\right) = \exp(-M u_1 / \lambda_0). \end{aligned} \quad (32)$$

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  $u_1 M e^{-\lambda_0 s}$  at time  $T_M - s$ , and by (12) will not die out by time  $T_M$  with probability  $\lambda_1 / (a_1 - b_1 e^{-\lambda_1 s})$ . Thinking about thinning a Poisson process we see that the number of mutations to type 1 that survive to time  $T_M$  is Poisson with mean

$$\mu(M) = M u_1 \int_0^\infty e^{-\lambda_0 s} \frac{\lambda_1}{a_1 - b_1 e^{-\lambda_1 s}} ds, \quad (33)$$

and it follows that

$$P(Z_1(T_M) > 0 | \Omega_\infty^0) = 1 - \exp(-\mu(M)). \quad (34)$$

The integral in (33) 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

$$\mu(M) = \frac{Mu_1}{\lambda_0} \int_0^1 \frac{\lambda_1}{a_1 - b_1 t^{\lambda_1/\lambda_0}} dt. \quad (35)$$

To compare with (7) in [24], we change notation

$$\begin{array}{cccccc} \text{here} & a_0 & b_0 & a_1 & b_1 & u_1 \\ [24] & r & d & a & b & ru \end{array}$$

To explain the last conversion recall that in [24] their mutation rate  $u$  is per cell division, while ours is a rate per unit time. The final conversion is that their  $\alpha = \lambda_1/\lambda_0$ , which in our notation is  $1/\alpha$ . For this reason, we will let  $\bar{\alpha} = \lambda_1/\lambda_0$ . Once these identifications are made, one can recognize their

$$F = \int_0^1 \frac{1 - b/a}{1 - (b/a)y^{\bar{\alpha}}} dy \quad (36)$$

as our integral and the two formulas are identical. In our notation the final result is

$$P(Z_1(T_M) > 0 | \Omega_\infty^0) = 1 - \exp\left(-\frac{Mu_1 F}{a_0 - b_0}\right). \quad (37)$$

When mutation does not change the birth or death rates, i.e.,  $a_0 = a_1 = a$  and  $b_0 = b_1 = b$ ,

$$\int_0^1 \frac{dt}{a - bt} = -\frac{1}{b} \log(a - bt) \Big|_0^1 = \frac{1}{b} \log\left(\frac{a}{a - b}\right),$$

and (37) becomes

$$P(Z_1(T_M) > 0 | \Omega_\infty^0) = 1 - \exp\left[-\frac{Mu_1 a}{b} \log\left(\frac{a}{a - b}\right)\right]. \quad (38)$$

Recalling  $u_1 = au$ , this agrees with (8) in [24] and (1) in Tomasetti and Levy [38] (there  $a = L$ ,  $b = D$ ).

The derivation in [24] is clever but not 100% rigorous. They break things down according to the number of sensitive cells  $k$ . The number increases to  $k + 1$  at rate  $kr$  and decreases to  $k - 1$  at rate  $kd$ . This shows that the number of sensitive cells is a time change of a random walk. If we let  $f_x(t)$  be the probability there are  $x$  sensitive cells at time  $t$  then the expected number of resistant cancer cells produced when there are  $x$  sensitive cancer cells is

$$R_x = ru \int_0^\infty x f_x(t) dt.$$

We are interested in what happens when the sensitive are conditioned to not die out and the process is stopped when  $M$  is reached. In this case, if we ignore the values

near the endpoints 0 and  $M$  then the number of man-hours at a site,  $\int_0^\infty x f_x(t) dt$ , will be the same as the occupation time of  $x$  in a random walk with drift. At time  $t$  the random walk will have moved from 0 to  $\approx (r - d)t$  so the time spent at each value in  $[0, (r - d)t]$  is  $\approx 1/(r - d)$ . This gives

$$R_x \approx \frac{u}{1 - d/r} \quad (39)$$

in agreement with the result computed in their appendix B. The (minor) flaw in their argument is that the  $R_x$  are random with  $ER_x \approx u/(1 - d/r)$ , but successive values are correlated. However, the correlations are short-range, so weighted averages of the  $R_x$ , such as the ones that appear in sums approximating (35), will be close to their mean.

**Approximations.** 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 (35) is  $\leq 1$ . 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  $\lambda_1/a_1$ , so using the reasoning that led to (34)

$$P(Z_1(T_M) = 0 | \Omega_\infty^0) \leq P(\sigma_1 > T_M | \Omega_\infty^0) = \exp(-Mu_1\lambda_1/a_1\lambda_0), \quad (40)$$

an inequality which can also be derived by noting that  $\lambda_1/a_1$  is a lower bound on the integrand.

**Example 3.** Leder et al. (2011) have taken a closer look at the probability of pre-existing resistance in chronic myeloid leukemia in order to obtain estimates of the diversity of resistant cells. 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  $u = 0.008 \cdot 10^{-7} = 8 \times 10^{-10}$ .

They examined eleven BCR-ABL mutations that produced resistance to treatment by imatinib, dasatinib, and nilotinib. The mutants are listed in the next table in decreasing order of their birth rates (per day). They assumed that the death rate in all cases is 0.003. To explain the names, *T315I* has a Threonine instead of Isoleucine at position 315, while p210 is the wild type.



mutation	growth rate	resistant to
T351I	0.0088	all
E255K	0.0085	I
Y235F	0.0082	I
p210	0.0080	none
E255V	0.0078	I
V299L	0.0074	D
Y253H	0.0074	I
M351I	0.0072	I
F317L	0.0071	I,D
T315A	0.0070	D
F317V	0.0067	D
L248R	0.0061	I,D

The growth parameters for T351I 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

$$\begin{aligned}
P(\tau_1 \leq T_M | \Omega_\infty^0) &= 1 - e^{-0.016} = 0.01587 \\
P(Z_1(T_M) > 0 | \Omega_\infty) &= 0.01263 \\
P(\sigma_1 \leq T_M | \Omega_\infty) &= 1 - e^{-0.010544} = 0.01049
\end{aligned}$$

where the last answer comes from evaluating the integral in (33) numerically. The mutation with the lowest growth rate in the table is L248R, which changes Leucine to Arginine at position 248. It has growth parameters  $a_1 = 0.0061$  and  $b_1 = 0.003$ , so  $\lambda_1 = 0.0031$ . Again  $Mu_1/\lambda_0 = 0.0058$  but this time  $\lambda_1/a_1 = 0.581$ , so we have

$$\begin{aligned}
P(\tau_1 \leq T_M | \Omega_\infty^0) &= 1 - e^{-0.016} = 0.01587 \\
P(Z_1(T_M) > 0 | \Omega_\infty) &= 0.01198 \\
P(\sigma_1 \leq T_M | \Omega_\infty) &= 1 - e^{-0.008131} = 0.00810
\end{aligned}$$

Comparing these two extreme cases we see that for all 11 mutations the probability of resistance is between 0.01198 and 0.01263. From this they conclude that for the parameters used in the computation the probability of no resistant type is 0.87, while there will be one with probability 0.12, and two with probability 0.01.

## 7 Accumulation of neutral mutations

It is widely accepted that cancers result from an accumulation of mutations that increase the growth rates of tumor cells compared to the cells that surround them. A number of studies have sequenced the genomes of tumors in order to find the causative or “driver” mutations. [58, 67, 71, 74, 77]. The Cancer Genome Atlas project and the International Cancer Genome Consortium represent the largest of such efforts.

They aim to sequence hundreds of examples of dozens of tumor types. See [46] for a summary of some of the results that have been found. There is much more on the web pages of these two organizations.

The search for “driver” mutations that cause the disease has been complicated by the fact that a typical solid tumor has dozens of amino-acid altering substitutions but only a small fraction of these are involved in the disease. The others are “passenger” mutations that are genetically neutral. A recent study [75] argues that half or more of the somatic mutations found in tumors occur before the onset of disease. To begin to understand the behavior of neutral mutations in our cancer model, we first consider those that occur to type 0’s. As shown in Section 1.1, if we condition  $Z_0(t)$  on the event  $\Omega_\infty^0$  that it does not die out, and let  $Y_0(t)$  be the number of individuals at time  $t$  whose families do not die out, then  $Y_0(t)$  is a Yule process in which births occur at rate  $\lambda_0$ . Since each of the  $Z_0(t)$  individuals at time  $t$  has a probability  $\gamma = \lambda_0/a_0$  of starting a family that does not die out, and the events are independent for different individuals,

$$Y_0(t)/Z_0(t) \rightarrow \gamma \quad \text{in probability,} \quad (41)$$

As explained in Section 1.1,  $Y_0(t)$  is the backbone of the branching tree. The remainder of the population lives in finite trees attached to the backbone. The finite trees are copies of the branching process conditioned to die out, so most of these are small, i.e., each individual in the population has a close relative on the backbone. Because of this, it is enough to study the accumulation of mutations in  $Y_0(t)$ .

Our first problem is to investigate the population site frequency spectrum,

$$F(x) = \lim_{t \rightarrow \infty} F_t(x), \quad (42)$$

where  $F_t(x)$  is the expected number of neutral “passenger” mutations present in more than a fraction  $x$  of the individuals at time  $t$ . To begin to compute  $F(x)$ , we note that by remarks, it is enough to investigate the frequencies of neutral mutations within  $Y_0$ . If we take the viewpoint of the infinite alleles model, where each mutation is to a type not seen before, then results can be obtained from Durrett and Schweinsberg’s [15] study of a gene duplication model. In their system there is initially a single individual of type 1. No individual dies and each individual independently gives birth to a new individual at rate 1. When a new individual is born it has the same type as its parent with probability  $1 - r$  and with probability  $r$  is a new type which is different from all previously observed types.

Let  $T_N$  be the first time there are  $N$  individuals and let  $F_{S,N}$  be the number of families of size  $> S$  at time  $T_N$ . Omitting the precise error bounds given in Theorem 1.3 of [15], that result says

$$F_{S,N} \approx r\Gamma\left(\frac{2-r}{1-r}\right) NS^{-1/(1-r)} \quad \text{for } 1 \ll S \ll N^{1-r}. \quad (43)$$

The upper cutoff on  $S$  is needed for the result to hold. When  $S \gg N^{1-r}$ ,  $EF_{S,N}$  decays exponentially fast.

As mentioned above, the last conclusion gives a result for a branching process with mutations according to the infinite alleles model, a subject first investigated by Griffiths and Pakes [19]. To study DNA sequence data, we are more interested in the frequencies of individual mutations. Using ideas from Durrett and Schweinsberg [14] it is easy to show:

**Theorem 2.** *If passenger mutations occur at rate  $\nu$  then  $F(x) = \nu/\lambda_0 x$ .*

**Numerical example.** To illustrate the use of Theorem 2 suppose the time has been scaled so that  $a_0 = 1$ ,  $\lambda_0 = 0.01$  and  $\nu = 10^{-5}$ . In support of the numbers we note that Bozic et al. (2010) estimate that the selective advantage provided by a typical cancer driver mutation is  $0.004 \pm 0.0004$ . As for the second, if the per nucleotide mutation rate is  $10^{-9}$  and there are 1000 nucleotides in a gene then a mutation rate of  $10^{-6}$  per gene results. In this case Theorem 2 predicts if we focus only on one gene then the expected number of mutations with frequency  $> 0.1$  is

$$F(0.1) = 10^{-6+2+1} = 0.001 \tag{44}$$

so, to a good first approximation, no particular neutral mutation occurs with an appreciable frequency. Of course, if we are sequencing 20,000 genes then there will be a few dozen passenger mutations seen in a given individual. On the other hand there will be very few specific neutral mutations that will appear multiple times in the sample, i.e., there should be few false positives. However, the bad news for sequencing studies is that in many cancers there is a wide variety of driver mutations that accomplish the same end.

*Proof of Theorem 2.* We will drop the subscript 0 for convenience. For  $j \geq 1$  let  $T_j = \min\{t : Y_t = j\}$  and notice that  $T_1 = 0$ . Since the  $j$  individuals at time  $T_j$  start independent copies  $Y^1, \dots, Y^j$  of  $Y$ , we have

$$\lim_{s \rightarrow \infty} e^{-\lambda s} Y^i(s) = \xi_i,$$

where the  $\xi_i$  are independent exponential mean 1 (here time  $s$  in  $Y^i$  corresponds to time  $T_j + s$  in the original process). From the limit theorem for the  $Y^i$  we see that for  $j \geq 2$  the limiting fraction of the population descended from individual  $i$  at time  $T_j$  is

$$r_i = \xi_i / (\xi_1 + \dots + \xi_j), \quad 1 \leq i \leq j,$$

which as some of you know has a beta(1,  $j - 1$ ) distribution with density

$$(j - 1)(1 - x)^{j-2}.$$

For those who don't we give the simple proof of this fact. Note that

$$((\xi_1, \dots, \xi_j) | \xi_1 + \dots + \xi_j = t)$$

is uniform over all nonnegative vectors that sum to  $t$ , so  $(r_1, \dots, r_j)$  is uniformly distributed over the nonnegative vectors that sum to 1. Now the joint distribution of the  $r_i$  can be generated by letting  $U_1, \dots, U_{j-1}$  be uniform on  $[0, 1]$ ,  $U^{(1)} < U^{(2)} < \dots < U^{(j-1)}$  be the order statistics, and  $r_i = U^{(i)} - U^{(i-1)}$  where  $U^{(0)} = 0$  and  $U^{(j)} = 1$ . From this and symmetry, we see that

$$P(r_i > x) = P(r_j > x) = P(U_i < 1 - x \text{ for } 1 \leq i \leq j - 1) = (1 - x)^{j-1},$$

and differentiating gives the density.

If the neutral mutation rate is  $\nu$  then on  $[T_j, T_{j+1})$  mutations occur to individuals in  $Y$  at rate  $\nu j$ , while births occur at rate  $\lambda j$ , so the number of mutations  $N_j$  in this time interval has a shifted geometric distribution with success probability  $\lambda/(\lambda + \nu)$ , i.e.,

$$P(N_j = k) = \left( \frac{\nu}{\nu + \lambda} \right)^k \frac{\lambda}{\nu + \lambda} \quad \text{for } k = 0, 1, 2, \dots \quad (45)$$

The  $N_j$  are i.i.d. with mean

$$\frac{\nu + \lambda}{\lambda} - 1 = \frac{\nu}{\lambda}.$$

Thus the expected number of neutral mutations that are present at frequency larger than  $x$  is

$$\frac{\nu}{\gamma} \sum_{j=1}^{\infty} (1 - x)^{j-1} = \frac{\nu}{\lambda x}.$$

The  $j = 1$  term corresponds to mutations in  $[T_1, T_2)$  which will be present in the entire population.  $\square$

Theorem 2 describes the population site frequency spectrum. The next result describes the frequencies in a sample. Suppose we sample  $n$  individuals from the Yule process when it has size  $N\gamma$  and hence the whole population has size  $\sim N$ . Let  $\eta_{n,m}$  be the number of mutations present in  $m$  individuals.

**Theorem 3.** As  $N \rightarrow \infty$

$$E\eta_{n,m} \begin{cases} \rightarrow \frac{n\nu}{\lambda} \cdot \frac{1}{m(m-1)} & 2 \leq m < n, \\ \sim \frac{n\nu}{\lambda} \cdot \log(N\gamma) & m = 1. \end{cases} \quad (46)$$

where  $a_N \sim b_N$  means  $a_N/b_N \rightarrow 1$ .

To explain the result for  $m = 1$ , we note that, as Slatkin and Hudson (1991) observed, genealogies in exponentially growing population tend to be star-shaped, i.e., the most recent common ancestor of two individuals occurs near the beginning of the branching process. The time required for  $Y_0(t)$  to reach size  $N\gamma$  is  $\sim (1/\lambda) \log(N\gamma)$ , so the number of mutations on our  $n$  lineages is roughly  $n\nu$  times this. Note that, (i) for a fixed sample size,  $E\eta_{n,m}$ ,  $2 \leq m < n$  are bounded.

*Proof of Theorem 3.* We begin with a calculus fact that is easy for readers who can remember the definition of the beta distribution. The rest of us can simply integrate by parts.

**Lemma 2.** *If  $a$  and  $b$  are nonnegative integers*

$$\int_0^1 x^a(1-x)^b dx = \frac{a!b!}{(a+b+1)!}. \quad (47)$$

Differentiating the distribution function from Theorem 2 gives the density  $\nu/\lambda x^2$ . We have removed the atom at 1 since those mutations will be present in every individual and we are supposing the sample size  $n > m$  the number of times the mutation occurs in the sample. Conditioning on the frequency in the entire population, it follows that for  $m \leq 2 < n$  that

$$E\eta_{n,m} = \int_0^1 \frac{\nu}{\lambda x^2} \binom{n}{m} x^m(1-x)^{n-m} dx = \frac{n\nu}{\lambda m(m-1)},$$

where we have used  $n \ll N$  and the second step requires  $m \geq 2$ .

When  $m = 1$  the formula above gives  $E\eta_{n,1} = \infty$ . To get a finite answer we note that the expected number that are present at frequency larger than  $x$  is

$$\frac{\nu}{\lambda} \sum_{j=1}^{N\gamma} (1-x)^{j-1} = \frac{\nu}{\lambda x} (1 - (1-x)^{N\gamma}).$$

Differentiating (and multiplying by  $-1$ ) changes the density from  $\nu/\lambda x^2$  to

$$\frac{\nu}{\lambda} \left( \frac{1}{x^2} (1 - (1-x)^{N\gamma}) - \frac{1}{x} N\gamma (1-x)^{N\gamma-1} \right). \quad (48)$$

Ignoring the constant  $\nu/\lambda$  for the moment and noticing

$$\binom{n}{m} x^m(1-x)^{n-m} = nx(1-x)^{n-1}$$

when  $m = 1$  the contribution from the second term is

$$n \int_0^1 N\gamma (1-x)^{N\gamma+n-2} dx = n \cdot \frac{N\gamma}{N\gamma+n-1} < n$$

and this term can be ignored. Changing variables  $x = y/N\gamma$  the first integral is

$$\begin{aligned} & \int_0^1 \frac{1}{x} (1 - (1-x)^{N\gamma}) (1-x)^{n-1} dx \\ &= \int_0^{N\gamma} \frac{1}{y} (1 - (1-y/N\gamma)^{N\gamma}) (1-y/N\gamma)^{n-1} dy. \end{aligned}$$

To show that the above is  $\sim \log(N\gamma)$  we let  $K_N \rightarrow \infty$  slowly and divide the integral into three regions  $[0, K_N]$ ,  $[K_N, N\gamma/\log N]$ , and  $[N\gamma/\log N, N\gamma]$ . Outside the first interval,  $(1 - y/N\gamma)^{N\gamma} \rightarrow 0$  and outside the third,  $(1 - y/N\gamma)^{n-1} \rightarrow 1$  so we conclude that the above is

$$O(K_N) + \int_{K_N}^{N\gamma/\log N} \frac{1}{y} dy + O(\log \log N)$$

□

## 8 Properties of the gamma function

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

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt. \quad (49)$$

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

$$\Gamma(\alpha + 1) = \int_0^\infty t^\alpha e^{-t} dt = \int_0^\infty \alpha t^{\alpha-1} e^{-t} dt = \alpha \Gamma(\alpha). \quad (50)$$

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

$$\Gamma(\alpha)\Gamma(1-\alpha) = \frac{\pi}{\sin(\pi\alpha)}. \quad (51)$$

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

$$\Gamma(\alpha) \sim \frac{\pi}{\sin(\pi\alpha)} \sim \frac{1}{\alpha}, \quad (52)$$

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

## 9 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  $t \in (-\infty, \infty)$ , where the star is to remind us that we have extended  $Z_0$  to negative times. The constant  $V_0$  could be set equal to 1 by changing the origin of time. We will not do that because proving the result for a general  $V_0$  will make it easy to prove results for  $Z_k^*(t)$  by induction. The expected number of type 1 families that begin at negative times is  $V_0 \mu / \lambda_0$ . When  $V_0 = 1$  this is  $10^{-3}$  or smaller so the extension changes the behavior very little. However, as the reader will see, it makes the analysis much easier.

We begin by stating and discussing the main results before we get involved in the details of the proofs. Let  $\alpha = \lambda_0/\lambda_1$ ,

$$c_{\mu,1} = \frac{1}{a_1} \left( \frac{a_1}{\lambda_1} \right)^\alpha \Gamma(\alpha), \quad c_{h,1} = c_{\mu,1} \Gamma(1 - \alpha), \quad (53)$$

and  $c_{\theta,1} = c_{h,1}(a_0/\lambda_0)$ .

**Theorem 4.** *If we assume  $Z_0^*(t) = V_0 e^{\lambda_0 t}$  then as  $t \rightarrow \infty$ ,*

$$e^{-\lambda_1 t} Z_1^*(t) \rightarrow V_1.$$

(i)  $V_1$  is the sum of the points in a Poisson process with mean measure

$$\mu(x, \infty) = c_{\mu,1} u_1 V_0 x^{-\alpha}$$

(ii)  $E(e^{-\theta V_1} | V_0) = \exp(-c_{h,1} u_1 V_0 \theta^\alpha)$

(iii)  $P(V_1 > x | V_0) \sim x^{-\alpha} c_{h,1} u_1 V_0 / \Gamma(1 - \alpha) = c_{\mu,1} u_1 V_0 x^{-\alpha}$ .

(iv) If  $V_0$  is exponential( $\lambda_0/a_0$ ) then

$$E \exp(-\theta V_1) = (1 + c_{\theta,1} u_1 \theta^\alpha)^{-1}, \quad (54)$$

and (v)  $P(V_1 > x) \sim x^{-\alpha} c_{\theta,1} u_1 / \Gamma(1 - \alpha)$ .

Conclusion (ii) implies that the distribution of  $(V_1 | V_0)$  is a one-sided stable law with index  $\alpha$ . It is known that such distributions have a power law tail. Conclusion (iii) makes this precise. In words, when  $x$  is large the sum is bigger than  $x$  if and only if there is a point in  $(x, \infty)$  in the Poisson process. The result in (iii) was discovered by Iwasa, Nozawa, and Michor (2006) using simulation. The conclusions in (iv) follow from (ii) by taking expected value and using (16). Finally (v) follows from (iv) in the same way (iii) follows from (ii), by using a Tauberian theorem given in Lemma 4.

*Proof of (i).* Mutations to type 1 occur at times of a Poisson process with rate  $u_1 V_0 e^{\lambda_0 s}$ . Theorem 1 implies that 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

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

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 on  $(-\infty, \infty) \times (0, \infty)$  (we ignore the points with  $y_i = 0$ ) that has intensity

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

Here, one of the two factors of  $\lambda_1/a_1$  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  $\lambda_1/a_1$  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

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

which completes the proof of (i). □

*Proof of (ii).* 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 3.**  $E(e^{-\theta Z_1^*(t)} | V_0) = \exp\left(-u_1 \int_{-\infty}^t V_0 e^{\lambda_0 s} (1 - \tilde{\phi}_{1,t-s}(\theta)) ds\right)$ .

*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. □



Replacing  $\theta$  by  $\theta e^{-\lambda_1 t}$  in Lemma 3 and letting  $t \rightarrow \infty$

$$E(e^{-\theta V_1} | V_0) = \lim_{t \rightarrow \infty} \exp \left( -u_1 V_0 \int_{-\infty}^t e^{\lambda_0 s} (1 - \tilde{\phi}_{1,t-s}(\theta e^{-\lambda_1 t})) ds \right). \quad (55)$$

To calculate the limit, we note that by (15)

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

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

$$1 - \tilde{\phi}_{1,t-s}(\theta e^{-\lambda_1 t}) \rightarrow \frac{\lambda_1}{a_1} \int_0^\infty (1 - e^{-\theta x}) (\lambda_1/a_1) e^{\lambda_1 s} e^{-x e^{\lambda_1 s} \lambda_1/a_1} dx. \quad (56)$$

Using this in (55) and interchanging the order of integration

$$E(e^{-\theta V_1} | V_0) = \exp(-u_1 V_0 h(\theta)), \quad (57)$$

where

$$h(\theta) = \frac{\lambda_1^2}{a_1^2} \int_0^\infty (1 - e^{-\theta x}) \left[ \int_{-\infty}^\infty e^{\lambda_0 s} e^{\lambda_1 s} e^{-x e^{\lambda_1 s} \lambda_1/a_1} ds \right] dx. \quad (58)$$

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

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

Inserting this in (58) 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 (53) and using  $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ , see (50), gives

$$h(\theta) = c_{\mu,1} \int_0^\infty (1 - e^{-\theta x}) \alpha x^{-\alpha-1} dx. \quad (59)$$

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

$$h(\theta) = c_{\mu,1} \theta^\alpha \int_0^\infty e^{-y} y^{-\alpha} dy = c_{\mu,1} \Gamma(1-\alpha) \theta^\alpha = c_{h,1} \theta^\alpha, \quad (60)$$

which completes the proof of (ii).  $\square$

*Proof of (iii).* To show that  $V_1$  has a power law tail, we note that (57) and (60) imply that as  $\theta \rightarrow 0$ ,

$$1 - E(e^{-\theta V_1} | V_0) \sim c_{h,1} u_1 V_0 \theta^\alpha, \quad (61)$$

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 4.** *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 (61), it follows that  $(1 - \phi(\theta))/\theta \sim c_{h,1} u_1 V_0 \theta^{\alpha-1}$ . Apply Lemma 4 with  $\omega(\theta) = (1 - \phi(\theta))/\theta$ ,  $u(x) = 1 - F(x)$  which is decreasing and  $\rho = 1 - \alpha$  we conclude

$$1 - F(x) \sim \frac{c_{h,1} u_1 V_0}{\Gamma(1 - \alpha)} x^{-\lambda_0/\lambda_1},$$

which proves (iii) and completes the proof of Theorem 4.  $\square$

## 10 Moments of $Z_1(t)$

The fact that  $EV_1 = \infty$  is a signal of complications that we will now investigate. We return to considering the process  $Z_k(t)$  which starts with one type 0 at time 0. The first step is to compute expected values.  $EZ_0(s) = e^{\lambda_0 s}$  so, assuming  $\lambda_0 < \lambda_1$ ,

$$\begin{aligned} EZ_1(t) &= \int_0^t u_1 e^{\lambda_0 s} e^{\lambda_1(t-s)} ds = u_1 e^{\lambda_1 t} \int_0^t e^{-(\lambda_1 - \lambda_0)s} ds \\ &= e^{\lambda_1 t} \frac{u_1}{\lambda_1 - \lambda_0} (1 - e^{-(\lambda_1 - \lambda_0)t}). \end{aligned} \quad (62)$$

To investigate the limiting behavior of  $Z_1(t)$  we note that

$$M_t = e^{-\lambda_1 t} Z_1(t) - \int_0^t u_1 Z_0(s) e^{-\lambda_1(t-s)} ds \quad \text{is a martingale.}$$

The integral, call it  $I_t$ , is increasing in  $t$  and has expected value  $e^{-\lambda_1 t} E Z_1(t) \leq u_1(\lambda_1 - \lambda_0)$  so it converges to a limit  $I_\infty$ . Since  $M_t \geq -I_\infty$  the martingale convergence theorem implies that

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

Our next step is to show  $EW_1 = EI_\infty = u_1/(\lambda_1 - \lambda_0)$ . This follows from:

**Lemma 5.** For  $k \geq 0$ ,  $\sup_t E(e^{-\lambda_k t} Z_k(t))^2 < \infty$ .

*Proof.* The base case is easy. We look at the derivative  $\frac{d}{dt} E(e^{-\lambda_0 t} Z_0(t))^2$

$$\begin{aligned} &= -2\lambda_0 E(e^{-\lambda_0 t} Z_0(t))^2 + e^{-2\lambda_0 t} (E[a_0 Z_0(t)(2Z_0(t) + 1)] - E[b_0 Z_0(t)(2Z_0(t) - 1)]) \\ &= e^{-2\lambda_0 t} (a_0 + b_0) E Z_0(t) = e^{-\lambda_0 t} (a_0 + b_0), \end{aligned}$$

and it follows that  $\sup_t E(e^{-\lambda_0 t} Z_0(t))^2 < \infty$ . Next, we suppose  $\sup_t E(e^{-\lambda_{k-1} t} Z_{k-1}(t))^2 \leq c_{k-1} < \infty$  and consider the derivative  $\frac{d}{dt} E(e^{-\lambda_k t} Z_k(t))^2$

$$\begin{aligned} &= -2\lambda_k E(e^{-\lambda_k t} Z_k(t))^2 + e^{-2\lambda_k t} E[a_k Z_k(t)(2Z_k(t) + 1)] \\ &\quad - e^{-2\lambda_k t} E[b_k Z_k(t)(2Z_k(t) - 1)] + e^{-2\lambda_k t} E[u_k Z_{k-1}(t)(2Z_k(t) + 1)] \\ &= (a_k + b_k) e^{-2\lambda_k t} E Z_k(t) + u_k e^{-2\lambda_k t} E[Z_{k-1}(t)(2Z_k(t) + 1)]. \end{aligned}$$

To bound  $2u_k e^{-2\lambda_k t} E[Z_{k-1}(t)Z_k(t)]$ , we use the Cauchy-Schwarz inequality and  $y^{1/2} \leq 1 + y$  for  $y \geq 0$  to get

$$\begin{aligned} &\leq 2u_k e^{-(\lambda_k - \lambda_{k-1})t} E[e^{-2\lambda_{k-1} t} Z_{k-1}^2(t)]^{1/2} E[e^{-2\lambda_k t} Z_k^2(t)]^{1/2} \\ &\leq 2u_k e^{-(\lambda_k - \lambda_{k-1})t} c_{k-1}^{1/2} (1 + E[e^{-2\lambda_k t} Z_k^2(t)]). \end{aligned}$$

Comparison theorems for differential equations imply that  $E(e^{-\lambda_k t} Z_k(t))^2 \leq m(t)$  where  $m(t)$  is the solution of the differential equation

$$\frac{d}{dt} m(t) = a(t)m(t) + b(t), \quad m(0) = 0, \quad (63)$$

with  $a(t) = 2u_k c_{k-1}^{1/2} e^{-(\lambda_k - \lambda_{k-1})t}$ , and

$$b(t) = (a_k + b_k) e^{-2\lambda_k t} E Z_k(t) + 2u_k e^{-2\lambda_k t} E Z_{k-1}(t) + 2u_k c_{k-1}^{1/2} e^{-(\lambda_k - \lambda_{k-1})t}.$$

Solving (63) gives

$$m(t) = \int_0^t b(s) \exp\left(\int_s^t a(r) dr\right) ds.$$

Since  $a(t)$  and  $b(t)$  are both integrable,  $m(t)$  is bounded.  $\square$

At this point we have shown that

$$\begin{aligned} e^{-\lambda_1 t} Z_1(t) &\rightarrow W_1 \text{ with } EW_1 = u_1/(\lambda_1 - \lambda_0), \\ e^{-\lambda_1 t} Z_1^*(t) &\rightarrow V_1 \text{ with } EV_1 = \infty. \end{aligned}$$

To see the reason for the difference between the two limits, we will repeat part of the proof of the point process result (i) from Theorem 4 to  $Z_1(t)$ . 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

$$\int_0^\infty u_1 V_0 e^{\lambda_0 s} \frac{\lambda_1}{a_1} e^{-(\lambda_1/a_1)x e^{\lambda_1 s}} ds.$$

Changing variables  $t = (\lambda_1/a_1)x e^{\lambda_1 s}$  now leads to

$$\frac{u_1 V_0}{a_1} \left( \frac{\lambda_1 x}{a_1} \right)^{-\alpha} \int_{\lambda_1 x/a_1}^\infty t^{\alpha-1} e^{-t} dt, \quad (64)$$

while in the last section the result was

$$\frac{u_1 V_0}{a_1} \left( \frac{\lambda_1 x}{a_1} \right)^{-\alpha} \int_0^\infty t^{\alpha-1} e^{-t} dt.$$

When  $\lambda_1 x/a_1$  is large the integral in (64) is roughly the value of the integrand at the lower limit so the quantity above is  $\sim Cx^{-1} \exp(-\lambda_1 x/a_1)$  as  $x \rightarrow \infty$ .

The last calculation shows that  $W_1$  is roughly the infinite mean random variable  $V_1$  truncated at  $a_1/\lambda_1$ . However  $EV_0 = a_0/\lambda_0$  and in most of our applications  $u_1 a_0/a_1 \lambda_0$  is small so this truncation is much larger than the typical value of  $V_1$  and  $EW_1$  is not a good measure of the size of  $W_1$ .

## 11 Luria-Delbruck distributions

We now go back in time 70 years to the work of Luria and Delbruck (1943) on bacterial growth. Luria and Delbruck grew a number of bacterial populations until they reached a size of order  $10^8$  then exposed them to attack by bacteriophage (anti-bacterial virus). Only bacteria resistant to the phage survived. At the time there were two theories about the emergence of resistance. (i) Mutations from sensitive to resistant were constantly occurring even in the absence of the phage. (ii) A certain proportion of organisms were able to adapt to themselves to the new environment, and thereby acquired an immunity that is passed on to their offspring. Under the second scenario, one would expect a Poisson number of survivors, but that was not what was observed.

To make it easier to connect with published work, the normal cells will be called type 1 and mutants type 2. Type 1's will be assumed to grow deterministically. There

are two cases to consider (a) the type 2's grow deterministically or (b) the type 2's are a Yule process (no deaths). If type 2's were a branching process then we would have the model we have been studying. The point of this section is to see the analogous results in a simpler setting and to explore other techniques that have been used.

## 11.1 Deterministic growth of type 1's

To determine the distribution of the number of mutants under scenario (i), Luria and Delbruck defined a process that starts at time 0 with one normal cell and no mutants. They assumed that normal cells grow deterministically at rate  $\beta_1$  so

$$N(t) = e^{\beta_1 t}.$$

Resistant mutants appear at rate  $\mu N(t)$ , so the expected number of mutations at time  $t$  is

$$m(t) = \int_0^t \mu e^{\beta_1 s} ds = \frac{\mu}{\beta_1} (e^{\beta_1 t} - 1). \quad (65)$$

Mutants grow at rate  $\beta_2$ , so the expected number of mutants at time  $t$  is

$$EX(t) = \int_0^t \mu e^{\beta_1 s} e^{\beta_2(t-s)} ds = \begin{cases} \mu t e^{\beta_1 t} & \beta_1 = \beta_2, \\ \frac{\mu}{\beta_2 - \beta_1} (e^{\beta_2 t} - e^{\beta_1 t}) & \beta_1 \neq \beta_2. \end{cases} \quad (66)$$

In our cancer applications we have  $\beta_1 < \beta_2$  and the formula reduces to the one in (62). However, for the bacterial experiment, it is natural to assume  $\beta_1 = \beta_2$ , i.e., in the absence of phage the resistance mutation is neutral.

Crump and Hoel [5], analyzed the model using the theory of “filtered Poisson processes”:

$$X(t) = \sum_{i=1}^{M(t)} W_i(t - \tau_i),$$

where  $M(t)$  is the number of mutations by time  $t$ ,  $\tau_i$  is the time of the  $i$ th mutation, and in the current context  $W_i(s) = \exp(\beta_2 s)$ . Using equation (5.42) of Parzen [34], which is our Lemma 3 given in Section 9, the cumulant generating function

$$K(u, t) = \log E[e^{\theta X(t)}] = \int_0^t \mu e^{\beta_1 s} (\psi_{t-s}(\theta) - 1) ds, \quad (67)$$

where  $\psi_r(\theta) = E(e^{\theta W(r)})$ , which in this case is just  $\exp(\theta e^{\beta_1 r})$ .

Given a random variable  $Y$

$$\frac{d}{d\theta} \log E(e^{\theta Y}) = \frac{E(Y e^{\theta Y})}{E(e^{\theta Y})} = EY \quad \text{when } \theta = 0.$$

Differentiating again

$$\begin{aligned}\frac{d^2}{d\theta^2} \log E(e^{\theta Y}) &= \frac{E(Y^2 e^{\theta Y})E(e^{\theta Y}) - (EY e^{\theta Y})^2}{(E e^{\theta Y})^2} \\ &= E(Y^2) - (EY)^2 \quad \text{when } \theta = 0.\end{aligned}$$

In general,

$$\left. \frac{d^n}{d\theta^n} \log E(e^{\theta Y}) \right|_{\theta=0} = \kappa_n$$

where  $\kappa_n$  is the  $n$ th cumulant. Differentiating  $n$  times we have

$$\kappa_n(t) = \int_0^t \mu e^{\beta_1 s} E W^n(t-s) ds.$$

When  $n = 1$  this is the formula for the mean given in (65). When  $n = 2$ ,

$$\begin{aligned}\kappa_2(t) &= \int_0^t \mu e^{\beta_1 s} e^{2\beta_2(t-s)} ds \\ &= \begin{cases} \frac{\mu}{\beta_1 - 2\beta_2} (e^{\beta_1 t} - e^{2\beta_2 t}) & \beta_1 \neq 2\beta_2, \\ \mu t e^{2\beta_1 t} & \beta_1 = 2\beta_2. \end{cases}\end{aligned}$$

According to Zheng's (1999) survey [39], very little is known about the distribution of  $X(t)$ . However, the proof of Theorem 4 applies easily to this case. Since the type 1 process is deterministic, a mutation at time  $s$  contributes  $e^{-\beta_2 s}$  to  $e^{-\beta_2 t} X(t)$ . In order for this contribution to be  $> x$  we need  $s$  to be  $< -(1/\beta_2) \log x$  so using the formula in (65) the number is Poisson with mean

$$\frac{\mu}{\beta_1} (e^{-(\beta_1/\beta_2) \log x} - 1) = \frac{\mu}{\beta_1} (x^{-\beta_1/\beta_2} - 1) \quad \text{for } x \leq 1.$$

This is the Poisson process for our stable law with the points  $> 1$  removed.

## 11.2 Lea-Coulson (1949) formulation

As in the previous example, normal cells grow deterministically  $N(t) = e^{\beta_1 t}$  and mutants appear at rate  $\mu N(t)$ , but now each mutant starts a Yule process having birth rate  $\beta_2$ . We begin by deriving an equation for the generating function  $G(z, t) = E[z^{X(t)}]$ . Since jumps from  $k \rightarrow k+1$  occur at rate  $\beta_2 k + \mu e^{\beta_1 t}$  at time  $t$ ,

$$\frac{\partial G}{\partial t} = \sum_k (\beta_2 k + \mu e^{\beta_1 t}) (z^{k+1} - z^k) P(X(t) = k).$$

The second term  $= \mu e^{\beta_1 t} (z-1)G$ . To deal with the first one we rewrite it as

$$\beta_2 z(z-1) \sum_{k \geq 1} k z^{k-1} P(X(t) = k) = \beta_2 z(z-1) \frac{\partial G}{\partial z},$$

so we have

$$\frac{\partial G}{\partial t} = \beta_2 z(z-1) \frac{\partial G}{\partial z} + \mu e^{\beta_1 t} (z-1) G. \quad (68)$$

For our purposes it is more convenient to use the cumulant generating function  $K(\psi, t) = \log(E[e^{\psi X(t)}]) = \log G(e^\psi)$ . The chain rule tells us that

$$\begin{aligned} \frac{\partial K}{\partial t} &= \frac{1}{G(e^\psi)} \frac{\partial G}{\partial t}(e^\psi), \\ \frac{\partial K}{\partial \psi} &= \frac{1}{G(e^\psi)} \frac{\partial G}{\partial z}(e^\psi) e^\psi, \end{aligned}$$

so we have

$$\frac{\partial K}{\partial t} = \beta_2 (e^\psi - 1) \frac{\partial K}{\partial \psi} + \mu e^{\beta_1 t} (e^\psi - 1), \quad (69)$$

which agrees with (50) in Zheng [39] and pages 125–129 in Bailey [3]. Inserting  $K(\psi, t) = \sum_{j \geq 1} \kappa_j(t) \psi^j / j!$  into (69) and equating terms we have

$$\begin{aligned} \kappa_1'(t) &= \beta_2 \kappa_1(t) + \mu e^{\beta_1 t}, \\ \kappa_2'(t) &= \beta_2 \kappa_2(t) + 2\beta_2 \kappa_1(t) + \mu e^{\beta_1 t}. \end{aligned}$$

The formula for the mean is the same as in (66). When  $\beta_1 \neq \beta_2$  and  $\beta_1 \neq 2\beta_2$

$$\text{var}[X(t)] = \frac{\mu e^{\beta_2 t} [\beta_1 (1 + e^{\beta_1 - \beta_2 t} - 2e^{\beta_2 t}) + 2\beta_2 (e^{\beta_2 t} - 1)]}{(\beta_1 - \beta_2)(\beta_1 - 2\beta_2)}.$$

In the exceptional cases

$$\text{var}[X(t)] = \begin{cases} \frac{\mu}{\beta_2} e^{\beta_2 t} (1 - e^{\beta_2 t} + 2\beta_2 t e^{\beta_2 t}) & \beta_1 = 2\beta_2, \\ \frac{2\mu}{\beta_1} e^{\beta_1 t} (e^{\beta_1 t} - 1) - \mu t e^{\beta_1 t} & \beta_1 = \beta_2. \end{cases}$$

In the special case  $\beta_1 = \beta_2$  these go back to Bailey. Zheng claims credit for the general formulas, see his (52) and (53). Using this approach we could derive exact formulas for the variance of our process rather than simply the bounds on the second moment given in Section 10.

**Generating function.** If  $\beta_2 > \beta_1$ , then we can analyze the limiting behavior of  $e^{-\beta_2 t} X(t)$  as we did in Section 9. For this reason, we restrict our attention now to the case  $\beta_1 = \beta_2 = \beta$ , in which case (68) becomes

$$\frac{\partial G}{\partial t} = \beta z(z-1) \frac{\partial G}{\partial z} + \mu e^{\beta t} (z-1) G,$$

with initial condition  $G(z, 0) = 1$ . Alternatively, using (65) we can use the boundary condition

$$G(0, t) = \exp(-m(t)).$$

Changing variables  $\theta = (\mu/\beta)e^{\beta t}$ , so  $F(z, \theta) = G(z, \beta^{-1} \log(\beta\theta/\mu))$ , one arrives at

$$\frac{\partial F}{\partial \theta} = \frac{z(z-1)}{\theta} \frac{\partial F}{\partial z} + (z-1)F.$$

One solution with  $F(0, \theta) = e^{-\theta}$ , called the Lea-Coulson p.g.f. can be written as

$$F(z, \theta) = (1-z)^{\theta(1-z)/z} = \exp(\theta(f(z) - 1)), \quad (70)$$

$$\text{where } f(z) = 1 + \left(\frac{1-z}{z}\right) \log(1-z).$$

If we let  $\theta(t) = (\mu/\beta)e^{\beta t}$ , then in terms of  $G$  the boundary condition is

$$G(0, t) = F(0, \theta(t)) = e^{-\theta(t)}.$$

In the literature it is often remarked that this is not correct due to the value at  $t = 0$ . However, since  $\theta(t) = \int_{-\infty}^t \mu e^{\beta s} ds$  this corresponds to our process on  $(-\infty, \infty)$ .

Define the  $LD(\theta, \phi)$  distribution by its generating function

$$G(z, \theta, \phi) = \exp\left(\theta \left[\frac{1}{z} - 1\right] \log(1 - \phi z)\right) = (1 - \phi z)^{\theta(1-z)/z}. \quad (71)$$

If  $\theta(t) = (\mu/\beta)e^{\beta t}$  and  $\phi(t) = 1 - e^{-\beta t}$  then this is the exact generating function for  $X(t)$ , which first appeared as (30a) in Armitage [2] where it was attributed to Bartlett. If instead we take  $\phi(t) = 1$  it is the Lea-Coulson p.g.f.

We will now derive these results from Parzen's formula (67). Recalling the formula for the generating function of the Yule process given in (10)

$$\log E(z^{X(t)}) = - \int_0^t \mu e^{\beta s} \frac{1-z}{1-z+ze^{-\beta(t-s)}} ds.$$

Changing variables  $x = ze^{-\beta(t-s)}$  so that  $dz = \beta ze^{-\beta(t-s)} ds$  and  $e^{\beta s} ds = (e^{\beta t}/\beta z) dx$  the above

$$\begin{aligned} &= - \frac{\mu e^{\beta t}}{\beta z} \int_{ze^{-\beta t}}^z \frac{1-z}{1-z+x} dx \\ &= \frac{\mu e^{\beta t}}{\beta} \left[ \frac{1-z}{z} \right] \log(1-z+ze^{-\beta t}), \end{aligned}$$

so changing notation we have (71):

$$G(z, \theta, \phi) = \exp\left(\theta \left[\frac{1-z}{z}\right] \log(1 - \phi z)\right)$$

If we replace 0 by  $-\infty$  in the lower limit then the last term becomes  $\log(1-z)$  instead.



The Lea-Coulson model is a special case of our branching process conditioned on  $V_0$ , so Theorem 4 implies that when  $\beta_1 < \beta_2$  the tail of the distribution for  $V_1$  has

$$P(V_1 > x|V_0) \sim c_{\mu,1}u_1V_0x^{-\beta_1/\beta_2}.$$

From this it should not be surprising that if  $X = LD(\theta, 1)$  then

$$P(X > n) \sim \theta/n \quad \text{and} \quad P(X = n) \sim \theta/n^2.$$

A nice proof with references to the earlier contributions can be found in Zheng [41].

## 12 Number of type 1's at time $T_M$

In Section 6 we studied the probability that type 1 cells were present when the number of type 0's reached size  $M$ . Since the 1's may be tumor cells with more aggressive growth or resistant to therapy, it is also important to understand the number that are present.

As in Section 6, Iwasa, Nowak, and Michor [24] study  $Z_1(T_M)$  by decomposing according to the size of the process when mutations happen. Their formula (4) for the generating function of the number of resistant cancer cells (written here in our notation) is:

$$G(\xi) = \exp\left(-\frac{u}{1-b_0/a_0} \sum_{x=1}^{M-1} (1-g_x(\xi))\right) \quad (72)$$

Here  $g_x(\xi)$  is the generating function of the branching process  $\tilde{Z}_1(t)$  evaluated at  $t = (1/\lambda_0)\log(M/x)$ , which is the time the type 0's need to grow from size  $x$  to size  $M$ .

Here we will instead note that it is immediate from the proof of Lemma 3 that

$$E \exp(-\theta Z_1^*(T_M)) \approx \exp\left(-u_1 \int_{-\infty}^0 M e^{\lambda_0 s} (1 - \tilde{\phi}_{-s}(\theta)) ds\right) \quad (73)$$

where  $\tilde{\phi}$  is the Laplace transform of the distribution of  $\tilde{Z}_1(t)$ , the branching process starting from a single 1 (and no mutations from 0 to 1). To see this note that in the previous formula the mutation rate at time  $s$  was  $u_1 V_1 e^{\lambda_0 s}$  for  $s \leq t$ , while now it is  $u_1 M e^{\lambda_0 s}$  for  $s \leq 0$ .

From (73), it is immediate that

**Theorem 5.** *If  $M \rightarrow \infty$  and  $Mu_1 \rightarrow \nu \in (0, \infty)$ ,  $Z_1^*(T_M)$  converges in distribution to  $U_1$  with Laplace transform*

$$E(\exp(-\theta U_1)) = \exp\left(-\nu \int_0^\infty e^{-\lambda_0 t} (1 - \tilde{\phi}_t(\theta)) dt\right)$$

To explain the assumption that  $Mu_1 \rightarrow \nu \in (0, \infty)$ , note that this implies that resistance is neither certain nor impossible. More concretely, it is estimated that in chronic myeloid leukemia [32] that  $M = 2.5 \times 10^5$  while  $u = 4 \times 10^{-7}$  so  $Mu \approx 0.1$ .

Since (11) implies

$$1 - \tilde{\phi}_t(\theta) = \frac{\lambda_1(1 - e^{-\theta})}{a_1(1 - e^{-\theta}) - e^{-\lambda_1 t}(b_1 - a_1 e^{-\theta})}$$

the Laplace transform of  $U_1$  is not pretty. However, as we will now show  $U_1$  has a power law tail, a result that [24] demonstrated by simulation.

To do this we note that if there is a mutation before  $T_M - (1/\lambda_1) \log y$  then it is likely that  $\{U_1 > y\}$ . The expected number of such mutations is

$$Mu_1 \int_{-\infty}^{-(1/\lambda_1) \log y} e^{\lambda_0 s} ds = Mu_1 \cdot \frac{1}{\lambda_0} e^{-(\lambda_0/\lambda_1) \log y} = (1/\lambda_0) \nu y^{-\alpha}.$$

As in Section 9 one can prove this rigorously by looking at the asymptotics for the Laplace transform as  $\theta \rightarrow 0$ . Changing variables  $t = -t(\theta) + x$  where  $t(\theta) = (1/\lambda_1) \log(1 - e^{-\theta})$

$$\begin{aligned} & \int_0^\infty dt e^{-\lambda_0 t} \frac{\lambda_1(1 - e^{-\theta})}{a_1(1 - e^{-\theta}) - e^{-\lambda_1 t}(b_1 - a_1 e^{-\theta})} \\ &= (1 - e^{-\theta})^\alpha \int_{-t(\theta)}^\infty e^{-\lambda_0 x} \frac{\lambda_1}{a_1 - e^{-\lambda_1 x}(b_1 - a_1 e^{-\theta})} dx \\ &\sim \theta^\alpha \int_{-\infty}^\infty e^{-\lambda_0 x} \frac{\lambda_1}{a_1 - e^{-\lambda_1 x}(b_1 - a_1)} dx \quad \text{as } \theta \rightarrow 0. \end{aligned}$$

The probability of having a type 1 at time  $T_M$  is, by (37),

$$P = 1 - \exp\left(-\frac{Mu_1 F}{1 - b_0/a_0}\right)$$

where  $F$  is given in (36). Using (72), we see that the mean of the number of resistance cells conditional on there being at least 1 is

$$\bar{Y} = \frac{G'(1)}{P} = \frac{u}{P(1 - b_0/a_0)} \sum_{x=1}^{M-1} \left(\frac{M}{x}\right)^{\bar{\alpha}},$$

where using notation from Section 6,  $\bar{\alpha} = \lambda_1/\lambda_0 = 1/\alpha$ . As observed in [24] this formula is not accurate in the case of advantageous resistant mutations because the dominant contribution comes from mutations when the tumor is small.

When mutations do not change the birth or death rates, i.e.,  $a_0 = a_1 = a$  and  $b_0 = b_1 = b$  then  $\bar{\alpha} = 1$  and the answer becomes

$$\frac{Mu_1 \log M}{P(1 - b/a)}.$$

If we suppose in addition that  $Mu_1$  is small then (38) implies

$$\frac{Mu_1}{P} \approx \frac{Mu - 1}{1 - \exp\left[-\frac{Mu_1 a}{\log}\left(\frac{a}{a-b}\right)\right]} \approx (a/b) \log(a/(a-b))$$

and we have

$$\bar{Y} \approx \frac{\log M}{(a/b - 1) \log(a/(a-b))},$$

which is (12) in [24] and (2) in [38]. Again the  $\log M$  comes from the unlikely event of resistance mutations that occur when the tumor is small.

### 13 Growth of $Z_k(t)$

The formulas for the constants in the next limit theorem are ugly, but the proof is very easy. The arguments in Section 9 and induction give the desired result. Let  $\alpha_k = \lambda_{k-1}/\lambda_k$ . Generalizing (53) we define

$$c_{\mu,k} = \frac{1}{a_k} \left(\frac{a_k}{\lambda_k}\right)^{\alpha_k} \Gamma(\alpha_k) \quad c_{h,k} = \Gamma(1 - \alpha_k) c_{\mu,k} \quad (74)$$

Let  $\mu_1 = u_1$  and inductively define for  $k \geq 2$

$$c_{\theta,k} = c_{\theta,k-1} c_{h,k}^{\lambda_0/\lambda_{k-1}} \quad (75)$$

$$\mu_k = \mu_{k-1} u_k^{\lambda_0/\lambda_{k-1}} = \prod_{j=1}^k u_j^{\lambda_0/\lambda_{j-1}}. \quad (76)$$

Let  $\mathcal{F}_\infty^{k-1}$  be the  $\sigma$ -field generated by  $Z_j^*(t)$ ,  $j \leq k-1$ ,  $t \geq 0$ .

**Theorem 6.** *If  $Z_0^*(t) = V_0 e^{\lambda_0 t}$  for  $t \in (-\infty, \infty)$  then as  $t \rightarrow \infty$ .*

$$e^{-\lambda_k t} Z_k^*(t) \rightarrow V_k \quad a.s.$$

(i)  $(V_k | \mathcal{F}_\infty^{k-1})$  is the sum of the points in a Poisson process with mean measure

$$\mu(x, \infty) = c_{\mu,k} u_k V_{k-1} x^{-\alpha_k}.$$

(ii)  $E(e^{-\theta V_k} | \mathcal{F}_\infty^{k-1}) = \exp(-c_{h,k} u_k V_{k-1} \theta^{\alpha_k})$ .

(iii)  $P(V_k > x | \mathcal{F}_\infty^{k-1}) \sim x^{-\alpha_k} c_{\mu,k} u_k V_{k-1} / \Gamma(1 - \alpha_k)$

(iv) If  $V_0$  is exponential( $\lambda_0/a_0$ ) then

$$E e^{-\theta V_k} = (1 + c_{\theta,k} \mu_k \theta^{\lambda_0/\lambda_k})^{-1} \quad (77)$$

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

*Proof.* We will prove this by induction. When  $k = 1$ , this follows from Theorem 4. Suppose now that  $k \geq 2$ . Let  $\mathcal{F}_t^{k-1}$  be the  $\sigma$ -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  $\tilde{\phi}_{k,t}(\theta) = Ee^{-\theta\tilde{Z}_k(t)}$ . The reasoning that led to Lemma 3 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  $e^{\lambda_{k-1}s}V_{k-1}$ ,  $\theta$  by  $\theta e^{-\lambda_k t}$ , and letting  $t \rightarrow \infty$

$$E(e^{-\theta V_k} | \mathcal{F}_\infty^{k-1}) = \lim_{t \rightarrow \infty} \exp\left(-u_k V_{k-1} \int_{-\infty}^t e^{\lambda_{k-1}s}(1 - \tilde{\phi}_{k,t-s}(\theta e^{-\lambda_k t})) ds\right) \quad (78)$$

At this point the calculation is the same as the one in the proof of Theorem 4 with 1 and 0 replaced by  $k$  and  $k-1$  respectively, so it follows from the proofs of (ii) and (iii) that

$$E(e^{-\theta V_k} | \mathcal{F}_\infty^{k-1}) = \exp(-c_{h,k} u_k V_{k-1} \theta^{\alpha_k}) \quad (79)$$

$$P(V_k > x | \mathcal{F}_\infty^{k-1}) \sim x^{-\alpha_k} c_{\mu,k} u_k V_{k-1} / \Gamma(1 - \alpha_k). \quad (80)$$

Taking expected value of (79) and using the result for  $k-1$

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

by (75) and (76), which proves (77). Part (v) now follows from Lemma 4.  $\square$

## 14 Transitions between waves

While the formulas in Theorem 6 are complicated, there is a simple underlying conceptual picture. For simplicity, consider the special case in which all the  $u_i = u$  and let  $L = \log(1/u)$ .

**Theorem 7.** *Let  $\beta_k = \sum_{j=0}^{k-1} 1/\lambda_j$ . 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 type  $k$  population grows like  $e^{\lambda_k t}$ , i.e., its logarithm grows like  $\lambda_k t$  and hence

$$\beta_{k+1} - \beta_k = \frac{1}{\lambda_k}. \quad (81)$$

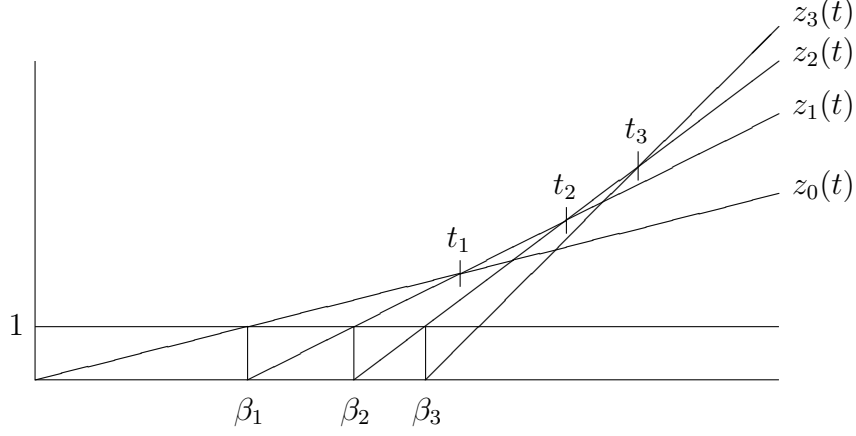


Figure 3: Growth of  $(1/L) \log^+(Z_k(Lt))$  for  $L = \log(1/u)$  shows the dominant type in the population as a function of  $t$ .

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

Theorem 7 makes it easy to obtain results for 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. The type  $k$ 's overtake the type  $k-1$ 's 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 (81)

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

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

**Theorem 8.** *If  $u_j \equiv u$  and  $\lambda_k = \lambda_0 + kb$  then  $T_k/L \rightarrow \beta_k + b^{-1}$*

## 15 Time to the first type $k$ , $k \geq 2$

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

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

Taking expected value and using Theorem 6

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

Using the definition of  $\mu_{k+1}$  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

$$t_{1/2}^{k+1} = \frac{1}{\lambda_0} \log \left( \frac{\lambda_k^{\lambda_0 / \lambda_k}}{c_{\theta,k} \mu_{k+1}} \right) = \frac{1}{\lambda_k} \log(\lambda_k) - \frac{1}{\lambda_0} \log(c_{\theta,k} \mu_{k+1}). \quad (82)$$

As in the case of  $\tau_1$ ,

$$P(\tau_{k+1} > t_{1/2}^{k+1} + x / \lambda_0 | \Omega_\infty^0) \approx (1 + e^x)^{-1}. \quad (83)$$

Again the result for the median  $s_{1/2}^{k+1}$  of the time  $\sigma_{k+1}$  of the first mutation to type  $k+1$  with a family that does not die out can be found by replacing  $u_{k+1}$  by  $u_{k+1} \lambda_{k+1} / a_{k+1}$ . Using  $\mu_{k+1} = \mu_k u_{k+1}^{\lambda_0 / \lambda_k}$  from (76), when we do this gives

$$s_{1/2}^{k+1} = \frac{1}{\lambda_k} \log \left( \frac{\lambda_k a_{k+1}}{u_{k+1} \lambda_{k+1}} \right) - \frac{1}{\lambda_0} \log(c_{\theta,k} \mu_k). \quad (84)$$

As in the case of  $\tau_{k+1}$ ,

$$P(\sigma_{k+1} > t_{1/2}^{k+1} + x / \lambda_0 | \Omega_\infty^0) \approx (1 + e^x)^{-1}. \quad (85)$$

To simplify and to relate our result to (S5) of Bozic et al (2010) we will look at the difference

$$s_{1/2}^{k+1} - s_{1/2}^k = \frac{1}{\lambda_k} \log \left( \frac{\lambda_k a_{k+1}}{u_{k+1} \lambda_{k+1}} \right) - \frac{1}{\lambda_{k-1}} \log \left( \frac{\lambda_{k-1} a_k}{u_k \lambda_k} \right) - \frac{1}{\lambda_0} \log \left( c_{h,k}^{\lambda_0 / \lambda_{k-1}} u_k^{\lambda_0 / \lambda_{k-1}} \right),$$

where in the second term we have used (75) and (76) to evaluate  $c_{\theta,k} / c_{\theta,k-1}$  and  $\mu_k / \mu_{k-1}$ . Recalling the formula

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

given in (74) we have

$$s_{1/2}^{k+1} - s_{1/2}^k = \frac{1}{\lambda_k} \log \left( \frac{\lambda_k^2 a_{k+1}}{a_k u_{k+1} \lambda_{k+1}} \right) - \frac{1}{\lambda_{k-1}} \log(\alpha_k \Gamma(\alpha_k) \Gamma(1 - \alpha_k)) \quad (86)$$

## 15.1 Relationship to Bozic et al. (2010)

Bozic et al. [4] investigated  $\sigma_k$  in order to obtain insights into the accumulation of passenger mutations. Their model takes place in discrete time, which facilitates simulation, their types are numbered starting from 1 rather than from 0. 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 cell can acquire an additional driver mutation with probability  $u$ . It is unfortunate that their birth probability  $b_j$  is our death rate for type  $j$  cells. We will not resolve this conflict because but we want to preserve their notation in order to make it easy to compare with the results in the paper.

They use  $\tau_j$  to denote  $\sigma_{j+1} - \sigma_j$ . In (S5) in their supplementary materials.

$$E(\sigma_{j+1} - \sigma_j) = \frac{T \log \left[ \frac{1-q_j}{ub_j(1-q_{j+1})} \left( 1 - \frac{1}{b_j(2-u)} \right) \right]}{\log[b_j(2-u)]}, \quad (87)$$

where  $q_j$  is probability that a type  $j$  mutation dies out. In quoting their result, we have dropped the  $1+$  inside the log in their formula, since it disappears in their later calculations and this makes their result easier to relate to ours.

When the differences in notation are taken into account (29) agrees with the  $j = 1$  case of (87). The death and birth probabilities in the model of Bozic et al. (2010) are  $d_1 = (1 - s)/2$  and  $b_1 = 1 - d_1 = (1 + s)/2$ , so  $\log(2b_1) \approx \log(1 + s) \approx s$ .  $q_j \approx (1 - js)/(1 + js) \approx 1 - 2js$ . Taking into account the fact that mutations occur only in the new daughter cell at birth, we have  $u_1 = b_1u$ , so when  $j = 1$  (87) becomes

$$E(\sigma_2 - \sigma_1) \approx \frac{1}{s} \log \left( \frac{s^2}{u_1 \cdot 2s} \right).$$

Setting  $\lambda_j = (j + 1)s$ , and  $a_i = b_{i+1}$  in our continuous time branching process, we have  $a_1/a_0 \approx 1$  and this agrees with (29).

**Example 4.** To match a choice of parameters studied in Bozic et al. (2010), we will take  $u = 10^{-5}$  and  $s = 0.01$ , so  $u_i = b_iu \approx 5 \times 10^{-6}$ , and

$$s_{1/2}^1 \approx \frac{1}{0.01} \log \left( \frac{10^{-4}}{5 \times 10^{-6} \cdot 0.02} \right) = 100 \log(1000) = 690.77.$$

Note that by (31) the fluctuations in  $\sigma_1$  are of order  $1/\lambda_0 = 100$ .

To connect with reality, we note that for colon cancer the average time between cell divisions is  $T = 4$  days, so 690.77 translates into 7.57 years. In contrast, Bozic et al. (2010) compute a waiting time of 8.3 years on page 18546. This difference is due to the fact that the formula they use ((1) on the cited page) employs the approximation  $1/2 \approx 1$ .

Turning to the later waves, we note that:

(i) the first “main” term in (86) corresponds to the answer in (87).

(ii) by (51),  $\alpha_k \Gamma(\alpha_k) \Gamma(1 - \alpha_k) = \pi \alpha_k / \sin(\pi \alpha_k) > 1$ , so the “correction” term not present in (87) is  $< 0$ , which is consistent with the fact that the heuristic leading to (87) considers only the first successful mutation.

To obtain some insight into the relative sizes of the “main” and the “correction” terms in (86), we will consider our concrete example in which  $\lambda_i = (i + 1)s$  and  $a_i = b_{i+1} \approx 1/2$ , so for  $i \geq 1$

$$s_{1/2}^{i+1} - s_{1/2}^i = \frac{1}{(i+1)s} \log \left( \frac{(i+1)^2 s}{u_{i+1}(i+2)} \right) - \frac{1}{is} \log \left( \frac{\pi \alpha_i}{\sin(\pi \alpha_i)} \right).$$

Taking  $s = 0.01$ ,  $u = 10^{-5}$  and  $u_i = 5 \times 10^{-6}$  leads to the results given in Table 2.

	main	corr.		from (86)	from (87)
$s_{1/2}^1$	690.77	0	$s_{1/2}^1$	690.77 (7.57)	550.87 (6.04)
$s_{1/2}^2 - s_{1/2}^1$	394.41	45.15	$s_{1/2}^2$	1040.03 (11.39)	895.39 (9.81)
$s_{1/2}^3 - s_{1/2}^2$	280.36	44.15	$s_{1/2}^3$	1276.24 (13.98)	1149.79 (12.60)

Table 1: Comparison of expected waiting times from (86) and (87). The numbers in parentheses are the answers converted into years using  $T = 4$  as the average number of days between cell divisions.

The values in the last column differ from the sum of the values in the first column because Bozic et al. (2010) indulge in some dubious arithmetic to go from their formula

$$E(\sigma_{j+1} - \sigma_j) = \frac{1}{js} \log \left( \frac{2j^2 s}{(j+1)u} \right),$$

to their final result

$$E\sigma_k \approx \frac{1}{2s} \log \left( \frac{4ks^2}{u^2} \right) \log k.$$

First they use the approximation  $j/(j+1) \approx 1$  and then  $\sum_{j=1}^{k-1} \approx \int_0^k$ . In the first row of the table this means that their formula underestimates the right answer by 20%. Bozic et al. (2010) tout the excellent agreement between their formula and simulations given in their Figure S2. However, a closer look at the graph reveals that while their formula underestimates simulation results, our answers agree with them almost exactly.



## 16 Application: Metastasis

Haeno, Iwasa and Michor [21] and Haeno and Michor [22] have used multitype branching processes to study metastasis, and have applied their work to study pancreatic cancer data [20]. Suppressing the complex details of the process of metastasis, the model has three types of cells

- Ordinary tumor cells, type 0.
- Cells that have the ability to metastasize, type 1.
- Cells that have spread to another location, type 2.

In the notation of [21] the birth rates of the three types of cells are  $r$ ,  $a_1$ , and  $a_2$ , while the death rates are  $d$ ,  $b_1$ , and  $b_2$ , so  $a_0 = r$  and  $b_0 = d$ . Their mutation rates, which we will call  $\mu_i$ , are per birth so in our terminology,  $u_i = a_{i-1}\mu_i$ .

The main questions are: what is the probability metastasis has occurred at diagnosis, and if so what is the size? To turn this into a precise mathematical question, we declare, as in Sections 6 and 12, the time of diagnosis to be  $T_M^0 = \inf\{t : Z_t^0 = M\}$ .

### 16.1 $P(Z_2(T_M) > 0)$

The first step is to calculate the probability of having a type 2 at time  $t$  given that we start with a single type 1 at time 0. This is related to the problem studied in Section 6 but here we do not condition on nonextinction of the 1's or use the approximation  $Z_1 \approx V_1 e^{\lambda_1 t}$ . Let  $g_i(z_0, z_1, t)$  be the generating function for  $(Z_1(t), Z_2(t))$  when the system is started from one individual of type  $i$ . Note that

$$P(Z_2(t) > 0 | Z_1(0) = 1) = 1 - g_1(1, 0, t).$$

To make it easier to derive the next result, we will think about the version of the model in which type 1's give birth at rate  $a_1$  and the result is type 2 with probability  $\mu_2$ , where  $\mu_2 = u_2/a_1$ . By considering what happens on the first jump, and arguing as in the proof of Lemma 1:

$$\begin{aligned} \frac{\partial g_1}{\partial t} &= b_1(1 - g_1) + a_1(1 - \mu_2)(g_1^2 - g_1) + a_1\mu_2(g_1g_2 - g_1), \\ \frac{\partial g_2}{\partial t} &= b_2(1 - g_2) + a_2(g_2^2 - g_2). \end{aligned} \tag{88}$$

These equations can be solved exactly, see Antal and Krapivsky [1], but their result, which involves hypergeometric functions, is not particularly useful.

If we let  $g(t) = g_1(1, 0, t)$  and  $h(t) = g_2(1, 0, t)$  then taking  $x = 0$  in (11), adjusting the indices, and multiplying top and bottom of the fraction by  $-1/a_2$

$$h(t) = 1 - \frac{1 - b_2/a_2}{1 - (b_2/a_2)e^{-(a_2 - b_2)t}}.$$

Our next goal is to show:

$$1 - g(t) \approx \frac{\lambda_1/a_1}{1 + (\lambda_1/a_1)^2(1/\mu_2)e^{-\lambda_1 t}}. \quad (89)$$

*Proof of (89).* Doing algebra on the first differential equation in (88) we have

$$\begin{aligned} \frac{dg}{dt} &= b_1(1 - g) + a_1(1 - \mu_2)(g^2 - g) + a_1\mu_2(gh - g) \\ &= (b_1 - a_1g)(1 - g) - a_1\mu_2g(g - h). \end{aligned} \quad (90)$$

Paraphrasing [21], if we neglect the second term, which is of order  $u_2$ , this equation is similar to a logistic equation and has two equilibria,  $g = 1$  and  $g = b_1/a_1$ . For small  $t$ ,  $g \approx 1$  and  $h \approx 0$  so the second term is approximately  $-a_1\mu_2$  and pushes the system away from the unstable equilibrium  $g = 1$  to the stable one at  $g = b_1/a_1$ . The second term is therefore only important for small  $t$  and we can approximate (90) as

$$\frac{dg}{dt} = (b_1 - a_1g)(1 - g) - a_1\mu_2 \frac{a_1g - b_1}{a_1 - b_1}.$$

The logic of this approximation is not explained in [21], but the final term is a linear function that is 0 when  $g = b_1/a_1$  and is 1 when  $g = 1$ . In addition, as the reader will soon see, it is convenient for computation. If we let  $\epsilon = a_1\mu_2/(a_1 - b_1)$  then the above can be written as

$$\frac{dg}{dt} = (b_1 - a_1g)(1 + \epsilon - g).$$

Using the calculation in the proof of (4) now with  $c = 1 + \epsilon$  we get from (6) that

$$g(t) = \frac{b_1 - (1 + \epsilon)e^{D(b_1 - (1 + \epsilon)a_1)}e^{(b_1 - (1 + \epsilon)a_1)t}}{a_1 - e^{D(b_1 - (1 + \epsilon)a_1)}e^{(b_1 - (1 + \epsilon)a_1)t}}.$$

Taking  $t = 0$  in (5) and recalling  $g(0) = 1$  we have

$$e^{D(b_1 - (1 + \epsilon)a_1)} = \frac{b_1 - a_1}{\epsilon}.$$

Plugging into the previous equation we have

$$g(t) = \frac{b_1 - (1 + \epsilon)(b_1 - a_1)\epsilon^{-1}e^{(b_1 - (1 + \epsilon)a_1)t}}{a_1 - (b_1 - a_1)\epsilon^{-1}e^{(b_1 - (1 + \epsilon)a_1)t}}.$$

Subtracting this from 1, and noting the second term in the numerator is  $(1 + \epsilon)$  times the second term in denominator we have

$$\begin{aligned} 1 - g(t) &= \frac{a_1 - b_1 + \epsilon(b_1 - a_1)\epsilon^{-1}e^{b_1 - (1 + \epsilon)a_1}}{a_1 - (b_1 - a_1)\epsilon^{-1}e^{(b_1 - (1 + \epsilon)a_1)t}} \\ &= \frac{a_1(1 + \epsilon) - b_1}{a_1 - (b_1 - a_1)\epsilon^{-1}e^{(b_1 - (1 + \epsilon)a_1)t}} - \epsilon \\ &= \frac{(\lambda_1/a_1) + \epsilon}{1 + (1 - b_1/a_1)\epsilon^{-1}e^{(b_1 - (1 + \epsilon)a_1)t}} - \epsilon. \end{aligned}$$

This does not match (A7) in [21], but remarkably when we simply we end up with the same end result. Plugging the definition of  $\epsilon = a_1\mu_2/(a_1 - b_1)$  into  $\epsilon^{-1}$  and replacing the other two  $\epsilon$  by 0 gives the desired result.  $\square$

As in the work of Iwasa, Nowak, and Michor [24] discussed in Section 6, [21] break things down according to the number of type 1 individuals that are present when the type 2 mutations occur to conclude that

$$P(Z_2(T_M^0) > 0) \approx 1 - \exp \left[ -\frac{\mu_1}{1 - b_0/a_0} \sum_{x=1}^M (1 - g_x^1(1, 0)) \right].$$

where  $g_x^1(s_1, s_2)$  is the bivariate generating function of  $(Z_1(t), Z_2(t))$  for a process started with one type 1 when  $Z_0(t) = x$ . As in (39) the first factor gives the expected number of type 1 mutations that occur when  $Z_0(t) = x$ . To evaluate the sum they assume deterministic growth of the type 0's to conclude that

$$g_x^1(1, 0) = g_1 \left( 1, 0, \frac{1}{\lambda_0} \log(M/x) \right) = g((1/\lambda_0) \log M).$$

Combining the last two formulas with (89), we have the formula derived in (A9) of [24]

$$P(Z_2(T_M^0) = 0) \approx \exp \left[ -\frac{\mu_1}{1 - b_0/a_0} \int_0^M \frac{\lambda_1/a_1 dx}{1 + (\lambda_1/a_1)^2 (1/\mu_2)(M/x)^{-\lambda_1/\lambda_0}} \right]. \quad (91)$$

**An alternative approach.** It follows from results given in Section 5 that

$$P(\sigma_2 > T_M^0) = \exp \left( - \int_{-\infty}^0 ds M e^{\lambda_0 s} u_1 \frac{\lambda_1}{a_1} P(\sigma_2 \leq s | \Omega_\infty^1) \right). \quad (92)$$

To see this note that  $u_1 \lambda_1/a_1$  is the rate for type 1 mutations that don't die out, so the integral gives  $\Lambda_{1,2}(M) =$  the mean number of successful type 1 mutations that produce a successful type 2 family before  $T_M^0$ . Since the number of such successes is Poisson with mean  $\Lambda_{1,2}(M)$ , the probability of none is  $\exp(-\Lambda_{1,2}(M))$ . Using (23)

$$P(\sigma_2 \leq t | \Omega_\infty^1) = 1 - (1 + c_{1,2} u_2 e^{\lambda_1 t})^{-1}. \quad (93)$$

where  $c_{1,2} = (a_1/\lambda_1^2) \lambda_2/a_2$ . To prepare for comparison with (91), we rewrite the right hand side as

$$\frac{c_{1,2} u_2 e^{\lambda_1 t}}{1 + c_{1,2} u_2 e^{\lambda_1 t}} = \frac{1}{1 + (1/u_2 c_{1,2}) e^{-\lambda_1 t}}. \quad (94)$$

Combining (92), (93), and (94), then changing variables  $x = M e^{\lambda_0 s}$ ,  $dx = \lambda_0 M e^{\lambda_0 s}$  we have

$$\begin{aligned} P(\sigma_2 > T_M^0) &\approx \exp \left[ -\frac{u_1}{\lambda_0} \int_0^M \frac{\lambda_1/a_1 dx}{1 + (1/u_2 c_{1,2})(M/x)^{-\lambda_1/\lambda_0}} \right] \\ &= \exp \left[ -\frac{u_1}{\lambda_0} \int_0^M \frac{\lambda_1/a_1 dx}{1 + (\lambda_1^2/a_1)(a_2/\lambda_2 u_2)(M/x)^{-\lambda_1/\lambda_0}} \right]. \end{aligned} \quad (95)$$

Since  $\mu_1 = u_1/a_0$  the factors in front are the same. In the term after the 1+ in denominator of the integral in (95),  $a_2/u_2 = \mu_2$ , so there is a factor of  $a_1/\lambda_2$  that separates this part of the formula from the corresponding part of (91).

## 16.2 Neutral case

Suppose that  $a_i \equiv a$ ,  $b_i \equiv b$ , and  $\lambda_i \equiv \lambda$ . In words, the mutation that confers the ability to migrate does not change the growth rate of the cancer cells. In this case,  $c_{1,2} = 1/\lambda$  so (95) becomes

$$P(\sigma_2 > T_M^0) \approx \exp \left[ -\frac{u_1}{a} \int_0^M \frac{dx}{1 + (\lambda/u_2)(x/M)} \right].$$

The integral is

$$\frac{u_2 M}{\lambda} \log(1 + (\lambda/u_2)(x/M)) \Big|_0^M,$$

so we have

$$P(\sigma_2 > T_M^0) \approx \exp \left( -\frac{M u_1 u_2}{\lambda a} \log(1 + \lambda/u_2) \right). \quad (96)$$

Since  $\lambda/u_2$  is large we can drop the 1+ inside the logarithm. If the quantity inside the exponent is small then we have

$$P(\sigma_2 \leq T_M^0) \approx \frac{M u_1 u_2}{\lambda a} \log(\lambda/u_2). \quad (97)$$

Using  $u_i = a\mu_i$ , and then dropping the  $\lambda/a$  from inside the logarithm as is done in [21]) the above becomes

$$P(\sigma_2 \leq T_M^0) \approx \frac{M \mu_1 \mu_2}{(\lambda/a)} \log(1/u_2),$$

which is  $\lambda/a$  times (5) in [21].

**Concrete Example.** The following situation (described in our notation) is simulated in panel (a) of Figure 2 in [21]:  $a_i = 0.2$ ,  $b_i = 0.1$ ,  $\lambda_i = 0.1$ ,  $u_1 = 2 \times 10^{-4}$ ,  $u_2 = 2 \times 10^{-6}$ , and  $M = 10^6$ .  $\log(10^4) = 9.210$  so the approximation from (97) is 0.184 versus 0.368 from (5). Despite the fact that we are estimating  $\sigma_2 < T_M^0$ , the first estimate is much closer to the data point for  $\mu_1 = 10^{-3}$ . Even closer is the one from (96), which is  $1 - \exp(-0.184) = 0.168$ .

## 17 Application: Ovarian cancer

This section summarizes results from [6]. Ovarian cancer is the fifth leading cause of cancer death among women in the United States with one in 71 American women

developing the disease during her lifetime. In 2012, 22,280 new cases are estimated to develop in the United States with 15,500 deaths expected [70]. To motivate our model, we begin by describing the four general stages used to classify the disease clinically:

- I. Cancer confined to one ovary
- II. Cancer involves both ovaries or has spread to other tissue within the pelvis
- III. Cancer has spread to the abdomen
- IV. Cancer has spread to distant organs

According to the SEER database [73], the distribution of the stage at diagnosis is (roughly) I: 20%, II: 10%, III: 40%, and IV: 30%. Five year survival statistics based on stage at diagnosis are as follows: I: 90%, II: 65%, III: 25%, IV: 10%, [47]. Given these statistics, the ability to accurately detect early stage disease could improve ovarian cancer survival dramatically. However, no screening strategy has yet been proven to reduce mortality [45]. Our goal is to estimate the size of the window of opportunity for screening, i.e., the amount of time in which screening can help improve survival. More specifically, it is the amount of time during which the primary tumor is of a size detectable by transvaginal ultrasound while the amount of metastasis has not significantly increased the chance of mortality.

Ovarian carcinoma begins as a tumor on the surface of the ovary or fallopian tube, which we call the primary tumor. Metastasis occurs either by direct extension from the primary tumor to neighboring organs, such as the bladder or colon, or when cancer cells detach from the surface of the primary tumor via an epithelial-to-mesenchymal (EMT) transition. Once the cells have detached, they float in the peritoneal fluid as single cells or multicellular spheroids. Cells then reattach to the omentum and peritoneum and begin more aggressive metastatic growth [59, 63]. We can therefore think of ovarian cancer as consisting of three general tumor cell subtypes:

- Primary (cells in the ovary or fallopian tube), type 0.
- Peritoneal (viable cells in peritoneal fluid), type 1.
- Metastatic (cells implanted on other intra-abdominal surfaces), type 2.

To parametrize the branching process model, we use data from [44], who examined the incidence of unsuspected ovarian cancers in apparently healthy women who underwent prophylactic bilateral salpingo-oophorectomies. They estimated that ovarian cancers had two-phase exponential growth with  $\lambda_0 = (\log 2)/4$  and  $\lambda_2 = (\log 2)/2.5$  per month, i.e., in the early stage the doubling time is 4 months, while in the later stage it is 2.5 months. The growth rate  $\lambda_1$  cannot be estimated from this data, so we will take it to be 0 and thereby get an upper bound on the size of the window of opportunity for screening, which was described in words above and will be defined more precisely in a minute. There does not seem to be data to allow us to directly

estimate the migration rates  $u_1$  and  $u_2$ . Fortunately, only the product  $u_1 u_2$  appears in our answers. We will choose  $u_1 u_2 = 10^{-4}$  to achieve agreement with observed quantities such as the size of the primary tumor when stage III is reached.

Type 0 cells are a branching process that grows at exponential rate  $\lambda_0 > 0$ . We are not interested in the situation in which the type 0's die out, so we consider the  $Z_0$ 's conditioned on nonextinction  $\Omega_\infty^0 = \{Z_0(t) > 0 \text{ for all } t\}$ . In this case (15) tells us that

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

Time  $t$  represents the amount of time since the initial mutation that began the tumor. That event is not observable, so by shifting the origin of time we can

**Assume**  $Z_0(t) = e^{\lambda_0 t}$  to get rid of  $V_0$ .

Type 1's leave from the surface of the primary tumor at rate  $u_1$  times the surface area. Ignoring the constant that comes from the relationship between the surface area and volume of a sphere, and letting  $\gamma_1 = 2\lambda_0/3$ , the mean is

$$\begin{aligned} EZ_1(t) &= \int_0^t u_1 e^{\gamma_1 s} e^{\lambda_1(t-s)} ds \\ &= \frac{u_1}{\gamma_1 - \lambda_1} (e^{\gamma_1 t} - e^{\lambda_1 t}) \sim \left( \frac{u_1}{\gamma_1 - \lambda_1} \right) e^{\gamma_1 t}. \end{aligned} \quad (98)$$

Since type 1's are cells floating in the peritoneal fluid and have less access to nutrients, it is natural to assume that  $\lambda_1 < \gamma_1$ . To remove the unknown rate  $\lambda_1$  from our calculations, we will later set  $\lambda_1 = 0$ . For the moment, we will proceed without that assumption.

**Theorem 9.** *If  $\gamma_1 > \lambda_1 \geq 0$  then  $Z_1(t)/EZ_1(t) \rightarrow 1$  in probability as  $t \rightarrow \infty$ .*

Intuitively, this holds since the integral in (98) has its dominant contribution from times  $s$  near  $t$  when there are a lot of migrations. The result is easily shown by computing second moments. The proof is given in the Section 17.3.

At time  $s$ , mutations occur to type 2 cells at rate  $u_2(u_1/\gamma_1)e^{\gamma_1 s}$ , so we let

$$s_2 = \frac{1}{\gamma_1} \log \left( \frac{\gamma_1}{u_1 u_2} \right) \quad (99)$$

be the time at which the mutation rate is 1. The next result follows easily from the proof of Theorem 6, but has the advantage of having simpler constants.

**Theorem 10.** *If  $\lambda_2 > \gamma_1 > 0$  then  $e^{-\lambda_2(t-s_2)} Z_2(t) \rightarrow V_2$  where  $V_2$  is the sum of points in a Poisson process with mean measure  $\mu(x, \infty) = C_2 x^{-\alpha_2}$  where  $\alpha_2 = \gamma_1/\lambda_2$ ,*

$$C_2 = \frac{1}{a_2} \left( \frac{a_2}{\lambda_2} \right)^{\alpha_2} \Gamma(\alpha_2).$$

## 17.1 Window of opportunity for screening

In order to be able to compute the size of the window of time in which screening can be effective, we need to have precise definitions of its two endpoints. For the upper bound, we define the time at which the patient enters stage III as  $T_2 = \min\{t : Z_2(t) = 10^9\}$ , where we have used the often-quoted rule of thumb that  $10^9$  cells =  $1 \text{ cm}^3 = 1$  gram. For the lower bound, we focus on detection by transvaginal ultrasound, so we define  $T_0 = \min\{Z_0(t) = 6.5 \times 10^7\}$ , corresponding to a spherical tumor of diameter 0.5 cm. These definitions are based on somewhat crude estimates of detectability and “significant” metastasis. If the reader prefers different values, it is easy to recalculate the size of the window.

Using our growth rate parameters  $\lambda_0 = (\log 2)/4 = 0.1733$  and  $\lambda_2 = (\log 2)/2.5 = 0.2772$ , we set

$$e^{0.1733T_0} = 6.5 \times 10^7, \quad \text{which gives} \quad T_0 = \frac{1}{0.1733} \log(6.5 \times 10^7) = 103.8.$$

months or 8.65 years. This may seem to be a very long time, but the estimate is consistent with calculations done for other types of cancers.

To make a crude calculation of  $T_2$ , ignoring the randomness in the growth of the 2's, we note that by (99), mutations to type 2 occur at rate 1 at time

$$s_2 = \frac{1}{\gamma_1} \log \left( \frac{\gamma_1}{u_1 u_2} \right).$$

From this point, it will take roughly

$$\frac{1}{\lambda_2} \log(10^9) = 74.76 \text{ months}$$

for the 2's to grow to size  $10^9$ . If we let  $u_1 u_2 = 10^{-4}$  and note  $\gamma_1 = 0.1155$ , then

$$s_2 = \frac{1}{.1155} \log(1155) = 61.05,$$

so  $T_2 = 74.76 + 61.05 = 135.81$  months, and the window of opportunity is  $T_2 - T_0 = 32.01$  months or 2.67 years.

To take the randomness of  $Z_2$  into account, we use Theorem 10 to conclude that

$$Z_2(t) \approx e^{\lambda_2(t-s_2)} V_2.$$

Setting this equal to  $10^9$  and solving, we have

$$T_2 \approx s_2 + \frac{1}{\lambda_2} \log(10^9/V_2), \tag{100}$$

and hence

$$T_2 = 138.51 - \log(1/V_2).$$

The distribution of the window,  $T_2 - T_0$ , is shown in Figure 4. When one takes the correction into account,  $T_2 - T_0$  is between 27 and 34 months with high probability.

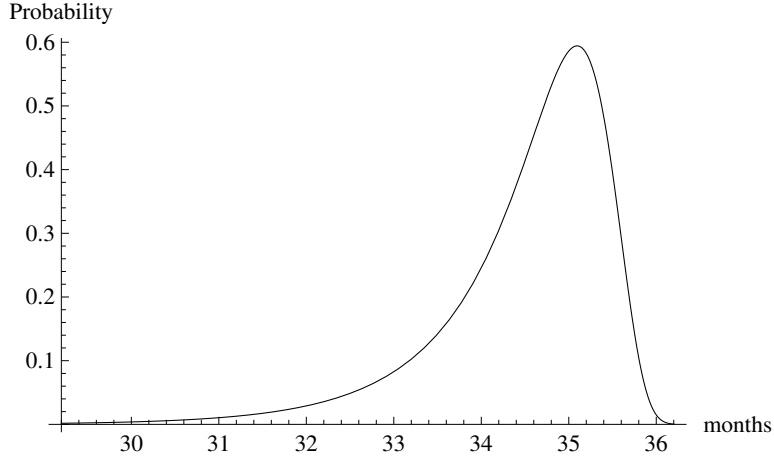


Figure 4: Distribution for  $T_2 - T_0$ . Due to a computational error in [6] the distribution needs to be shifted to the left by 2.7 months.

## 17.2 Primary tumor size at onset of metastasis

When tumors are found at an early stage and removed surgically, patients almost always undergo chemotherapy as the next step of treatment. Although chemotherapy is prescribed in part due to the possibility of residual tumor in the site of surgery, it is also done out of concern that some cells had already metastasized but had not yet grown to a detectable size. Here, we use our model to estimate the probability of metastasis given a primary tumor of known size.

We want to find the distribution for the size of the primary tumor at the onset of metastasis, i.e., at time  $T_2$ . Using (100), we have

$$Z_0(T_2) \approx \exp(\lambda_0 [s_2 + (1/\lambda_2) \log(10^9/V_2)]).$$

Using (99) and recalling  $\gamma_1 = 2\lambda_0/3$ ,  $\lambda_0/\lambda_2 = 5/8$ , we now have

$$Z_0(T_2) \approx \left(\frac{\gamma_1}{u_1 u_2}\right)^{3/2} \left(\frac{10^9}{V_2}\right)^{5/8}.$$

From this, we see that the size of the primary tumor at time  $s_2$  is

$$\left(\frac{\gamma_1}{u_1 u_2}\right)^{3/2} \approx 4 \times 10^4,$$

which translates into a diameter of 0.84 mm. If we ignore randomness and take  $V_2 = 1$ , then the size of the primary at time  $T_2$  is

$$4 \times 10^4 \times 10^{45/8} = 1.686 \times 10^{10},$$

which translates into 3 cm. Figure 5 shows the distribution for size and a comparison with the results of [44], respectively.



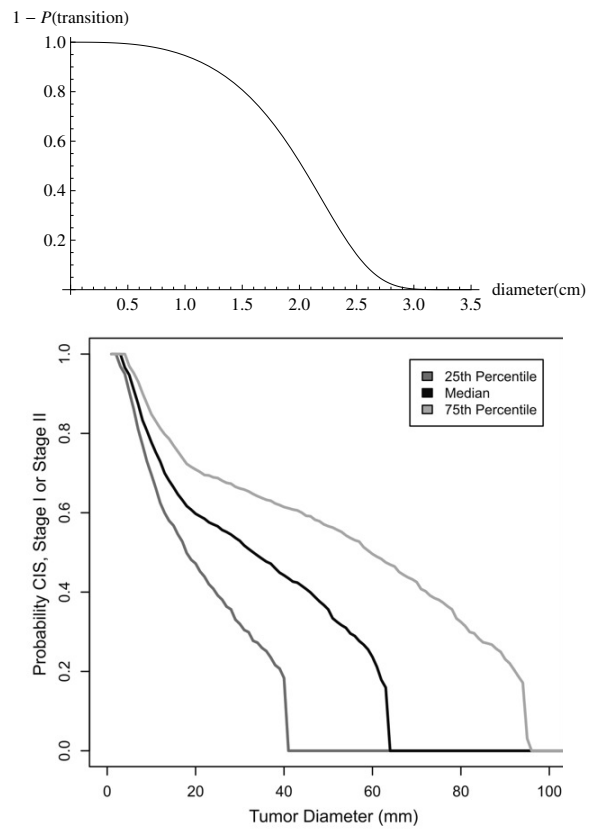


Figure 5: Computed distribution of the size of the primary at time of detection compared with data from [44].

There is an overall difference of factor two between the time at which the curves decrease to 0. This difference could be removed by adjusting our parameters. However, such an adjustment would not change the disagreement between the shapes of the early parts. We find the initial sharp drop of the curve surprising and think it might be an artifact of the way in which they analyzed data. In a typical Kaplan-Meier survival study, patients either leave the study or die. Deaths in such a study are observed when they occur. In the ovarian cancer study, in which a woman is observed to have stage III cancer, the progression occurred at some time in the past which must be estimated. The curve will be skewed if this is not done correctly.

### 17.3 Proof of Theorem 9

**Theorem 9.** If  $\gamma_1 > \lambda_1 \geq 0$  then  $Z_1(t)/EZ_1(t) \rightarrow 1$  in probability as  $t \rightarrow \infty$ .

*Proof.* Let  $\bar{Z}_1(t)$  be the contribution from mutations before time  $t - \log t$ .

$$E\bar{Z}_1(t) = \int_0^{t-\log t} u_1 e^{\gamma_1 s} e^{\lambda_1(t-s)} ds.$$

A little algebra and the asymptotic behavior of  $EZ_1(t)$  given in (98) shows that

$$E\bar{Z}_1(t) = e^{\lambda_1(\log t)} EZ_1(t - \log t) \sim e^{(\lambda_1 - \gamma_1)(\log t)} EZ_1(t).$$

Since  $\gamma_1 > \lambda_1$ ,  $E\bar{Z}_1(t)/EZ_1(t) \rightarrow 0$ .

Let  $Y_1(t)$  be the number of type 1's at time  $t$  when we start our multitype branching process with one 1 and no 0's. Using Theorem 1, there is a positive constant  $C$  so that

$$\text{var}(Y_1(t)) \sim Ce^{2\lambda_1 t}.$$

The process  $Z_0(t)$  is deterministic, so the contributions to  $Z_1(t)$  from different time intervals are independent. Thus, if we let  $\hat{Z}_1(t) = Z_1(t) - \bar{Z}_1(t)$ , we have

$$\text{var}(\hat{Z}_1(t)) \sim \int_{t-\log t}^t u_1 e^{\gamma_1 s} e^{2\lambda_1(t-s)} ds.$$

Introducing the normalization and changing variables  $r = t - s$  in the integral, we have

$$\begin{aligned} \text{var}(e^{-\gamma_1 t} \hat{Z}_1(t)) &\sim e^{-\gamma_1 t} \int_0^{\log t} e^{(2\lambda_1 - \gamma_1)r} dr \\ &\leq e^{-\gamma_1 t} (\log t) (1 + t^{2\lambda_1 - \gamma_1}) \rightarrow 0, \end{aligned}$$

where we have added 1 to take into account the possibility that  $2\lambda_1 - \gamma_1 < 0$ . Since  $e^{-\gamma_1 t} Z_1(t)$  converges to a positive limit, it follows that  $\text{var}(\hat{Z}_1(t)/EZ_1(t)) \rightarrow 0$ . Combining this with the fact that  $E\bar{Z}_1(t)/EZ_1(t) \rightarrow 0$  gives the desired result.  $\square$

## 18 Application: Intratumor heterogeneity

For several reasons problems in cancer treatment are caused by diversity of cell types in a tumor.

- Different subpopulations within a tumor and its metastases may have varying types of response to any given treatment. This has been documented by sequencing samples from different regions in renal carcinoma [54], glioblastoma [72], and breast cancer [64].
- Heterogeneity levels are associated with aggressiveness of disease. For example, in Barrett's esophagus [61] and in breast cancer [66]
- Heterogeneity has long been implicated in the development of resistance to cancer therapy after an initial response, and in the development of metastases [51]. This has important consequences for treatment [53, 68].

For simplicity, we will restrict our attention to quantifying the amount of diversity in the first wave. These results generalize easily to later waves. Since the population is (at most times) dominated by a single wave, this is enough to quantify the diversity in the entire population. More details about the generalization can be found in [12].

To formulate a simple mathematical problem, we will suppose that the type 1 producing mutations are all different, and the descendants of a mutant, which we will call a clone, have the same genotype. The point process in the proof of Theorem 4 allows us to understand the sizes of the clones. Recall that we are supposing  $Z_0(t) = V_0 e^{\lambda_0 t}$  where  $V_0$  is nonrandom. The results about diversity will not depend on  $V_0$  so this can be done without loss of generality.

As in Section 9, define a two dimensional point process  $\mathcal{X}_t$  with a point at  $(s, w)$  if there was a mutation to type 1 at time  $s$  and the resulting type 1 branching process  $\tilde{Z}_1(t)$  has  $e^{-\lambda_1(t-s)} \tilde{Z}_1(t) \rightarrow w$ . A point at  $(s, w)$  contributes  $e^{-\lambda_1 s} w$  to  $V_1 = \lim_{t \rightarrow \infty} e^{-\lambda_1 t} Z_1(t)$ .

$$V_1 = \sum_{(s,w) \in \mathcal{X}_t} e^{-\lambda_1 s} w$$

is the sum of points in a Poisson point process with mean measure  $\mu(z, \infty) = A_1 u_1 V_0 z^{-\alpha}$  where  $\alpha = \lambda_0 / \lambda_1$ .

In Section 9 we showed that the distribution of  $V_1$  is a one-sided stable law. To make the connection with the point process representation recall

**Theorem 11.** *Let  $Y_1, Y_2, \dots$  be independent and identically distributed nonnegative random variables with  $P(Y_i > x) \sim cx^{-\alpha}$  where  $0 < \alpha < 1$  and let  $S_n = Y_1 + \dots + Y_n$ . Then*

$$S_n / n^{1/\alpha} \rightarrow V,$$

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

*Ideas behind the proof.* The key is to let

$$N_n(z) = |\{1 \leq m \leq n : Y_m > zn^{1/\alpha}\}|.$$

The assumption  $P(Y_i > x) \sim cx^{-\alpha}$  implies that  $nP(Y_i > zn^{1/\alpha}) \sim cz^{-\alpha}/n$ . Since the variables are independent  $N_n(a)$  converges to a Poisson with mean  $cz^{-\alpha}$ . With a little work, one can show that the contribution of all the jumps  $\leq \epsilon n^{1/\alpha}$  is small if  $\epsilon$  is and the desired result follows. For more details see Section 3.7 in [9].  $\square$

**Simpson's index** is a useful measure of the amount of diversity. Called the homozygosity in genetics, it is defined to be the probability two randomly chosen individuals in wave 1 are descended from the same mutation.

$$R = \sum_{i=1}^{\infty} \frac{X_i^2}{V_1^2},$$

where  $X_1 > X_2 > \dots$  are points in the Poisson process and  $V_1$  is the sum. While the point process is somewhat complicated, the formula for the mean of Simpson's index is very simple.

**Theorem 12.**  $ER = 1 - \alpha$  where  $\alpha = \lambda_0/\lambda_1$ .

This was proved in [12] by using a 2001 result of Fuchs, Joffe and Teugels [18] about

$$R_n = \sum_{i=1}^n \frac{Y_i^2}{S_n^2} \quad \text{where} \quad S_n = \sum_{i=1}^n Y_i,$$

and the  $Y_i$  are the i.i.d. random variables from Theorem 11. They proved  $ER_n \rightarrow 1 - \alpha$  so the task in [12] was to show  $R_n \Rightarrow R$  and  $ER_n \rightarrow ER$ . See page 472 of [12] for more details. A more interesting derivation can be given using results about Poisson-Dirichlet distributions but that requires more machinery, so we postpone the proof until after we have stated all of our results

We can use a 1973 result of Logan, Mallows, Rice and Shepp [28] to get an idea about the distribution of  $R$ . Consider the "self-normalized sums"

$$S_n(p) = \frac{\sum_{i=1}^n Y_i}{(\sum_{j=1}^n Y_j^p)^{1/p}},$$

where  $Y_1, Y_2, \dots \geq 0$  are i.i.d. In the theory of sums of independent random variables this definition is nice because one does not have to know what the index of the stable law is in order to do the normalization. Of course, our motivation here is that

$$S_n(2) = R_n^{-1/2}.$$

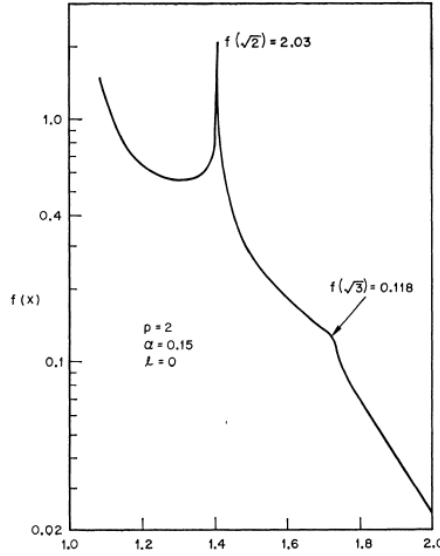


Figure 6: Picture of the limit distribution for  $S_n(2)$  when  $\alpha = 0.15$ .

Logan et al. [28] proved convergence in distribution and identified the Fourier transform of the limit. The convergence of

$$\sum_{i=1}^n Y_i/n^{1/\alpha} \quad \text{and} \quad \sum_{j=1}^n Y_j^p/n^{p/\alpha}$$

to stable laws are standard, but to get convergence of the ratio one needs to show convergence of the joint distribution. This is done with clever computations. Even in the relatively nice case  $p = 2$  there is not a good formula. The picture of the density function in the case  $p = 2$ ,  $\alpha = 0.15$  if Figure 6 might explain why. In the figure  $\ell$  is the asymmetry parameter of the stable law.  $\ell = 0$  means we are in the positive one sided case.

**Size of the largest clone.** Using a 1952 result of Darling [7], we can find the limiting distribution

$$M_n = \max_{1 \leq i \leq n} Y_i/S_n$$

of the fraction of individuals in the largest clone.

**Theorem 13.** *As  $n \rightarrow \infty$ ,  $1/M_n \rightarrow T$  where  $T$  has characteristic function  $e^{it}/f_\alpha(t)$  where*

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

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

## 18.1 Poisson-Dirichlet distributions

We now give a second derivation of Theorem 12. The proof is not simple, but it does take us through some interesting territory. The facts we use here, unless otherwise indicated can be found in the 1997 paper by Pitman and Yor [36]. We thank Jason Schweinsberg, Jim Pitman, and Ngoc Tran for helpful discussions of this material. This two parameter family of distributions  $PD(\alpha, \theta)$  can be defined from a *residual allocation model*. To do this we begin by recalling that a  $\text{beta}(\beta, \gamma)$  distribution has density on  $(0, 1)$  given by

$$\frac{\Gamma(\beta + \gamma)}{\Gamma(\beta)\Gamma(\gamma)} x^{\beta-1} (1-x)^{\gamma-1}. \quad (101)$$

Let  $B_1, B_2, \dots$  be independent with  $B_n \stackrel{d}{=} \text{beta}(1 - \alpha, \theta + n\alpha)$  distribution. Let  $Z_1 = B_1$  and for  $k \geq 2$  let

$$Z_k = (1 - B_1) \cdots (1 - B_{k-1}) B_k.$$

This sometimes called a stick breaking model since at time  $k$  we break off a fraction  $B_k$  of what remains of the stick.

Let  $U_1, U_2, \dots$  be the  $Z_i$  arranged in decreasing order. It is known that We can go from the  $U_i$  to a sequence  $\hat{U}_i$  with the same joint distribution as the  $Z_i$  by a *size-biased permutation*.

$$P(\hat{U}_{n+1} = U_j | \hat{U}_1, \dots, \hat{U}_n, U_1, U_2, \dots) = \frac{U_j}{1 - \hat{U}_1 \cdots \hat{U}_n},$$

if  $U_j$  has not already been chosen or 0 otherwise. An important corollary is

$$\hat{U}_1 \text{ has a } \text{beta}(1 - \alpha, \theta + \alpha) \text{ distribution.} \quad (102)$$

An example familiar from elementary probability is given by the limiting behavior of the cycle sizes of a random permutation  $\pi$  of  $\{1, 2, \dots, N\}$ . Construct the first cycle by following  $1 \rightarrow \pi(1) \rightarrow \pi(\pi(1)) \rightarrow \dots$  until we return to 1. When this happens the first cycle is complete. To continue we take the smallest value not in the first cycle and repeat the procedure. For more details see Example 2.2.4 in [9]. It is easy to see that if we let  $L_1, L_2, \dots$  be the cycle lengths and we take  $B_i$  to be  $\text{beta}(1, 1)$ , i.e., uniform on  $(0, 1)$ , then

$$(L_1/N, L_2/N, L_3/N, \dots) \Rightarrow (Z_1, Z_2, Z_3, \dots).$$

where  $\Rightarrow$  indicates convergence in distribution of the sequence, i.e., for each  $n$  the joint distribution of the first  $n$  terms converges.

A more sophisticated example appears in the limiting behavior of the Ewen's sampling formula for the infinite alleles model. As explained in Section 1.3.3 of [8],

if we let  $s_j(N)$  be the number of individuals with the  $j$ th most frequent allele when the rescaled mutation rate is  $\theta$ , and let the  $B_i$  be beta(1,  $\theta$ ) then

$$(s_1(N)/N, s_2(N)/N, \dots) \Rightarrow (U_1, U_2, \dots).$$

so the limit distribution is  $PD(0, \theta)$ . This example provided Kingman's motivation for introducing a one-parameter family of Poisson-Dirichlet distribution in [26].

The connection we are most interested in here comes from the following example. Let  $\tau_s$  be a one-sided stable process with index  $\alpha$ . That is,  $s \rightarrow \tau_s$  is nondecreasing, has independent increments and

$$E \exp(-\lambda(\tau_t - \tau_s)) = \exp\left(-\int_0^\infty (1 - \exp(-\lambda x)) (t - s) c x^{-\alpha} dx\right).$$

Let  $J_1(t) > J_2(t) > \dots$  be the jumps of  $\tau_s$  for  $s \in [0, t]$  listed in decreasing order. Proposition 6 in [36] implies that

$$\left(\frac{J_1(t)}{\tau_t}, \frac{J_2(t)}{\tau_t}, \dots\right) =_d PD(0, \alpha).$$

Changing to our notation this implies

$$\left(\frac{X_1}{V_1}, \frac{X_2}{V_1}, \dots\right) =_d PD(0, \alpha). \quad (103)$$

Returning to other notation introduced earlier let  $(U_1, U_2, \dots) =_d PD(\theta, \alpha)$  and  $\hat{U}_i$  be a size biased permutation. Formula (6) in [36] gives us

$$\begin{aligned} E_{\alpha, \theta} \sum_{n=1}^{\infty} f(U_n) &= E_{\alpha, \theta} \left[ \frac{f(\hat{U}_1)}{\hat{U}_1} \right] \\ &= \frac{\Gamma(\theta + 1)}{\Gamma(\theta + \alpha)\Gamma(1 - \alpha)} \int_0^1 du \frac{f(u)}{u} u^{-\alpha} (1 - u)^{\alpha + \theta - 1}. \end{aligned} \quad (104)$$

The first equality follows easily from the fact that  $\hat{U}_1 = U_n$  with probability  $U_n$ , the second from (102) and the formula for the beta density given in (101). If we let  $f(u) = u^2$  then we have that

$$\begin{aligned} E \sum_{n=1}^{\infty} \frac{X_n^2}{V_1^2} &= \frac{1}{\Gamma(\alpha)\Gamma(1 - \alpha)} \int_0^1 du u^{1-\alpha} (1 - u)^{\alpha-1} \\ &= \frac{1}{\Gamma(\alpha)\Gamma(1 - \alpha)} \cdot \Gamma(2 - \alpha)\Gamma(\alpha) = 1 - \alpha, \end{aligned} \quad (105)$$

by the Gamma function recursion (50).

A second commonly used measure of diversity is the entropy, which is often called Shannon's index in the literature. Taking  $f(u) = u \log u$  in (104), we see that the expected value of the entropy is

$$\frac{\Gamma(\theta + 1)}{\Gamma(\theta + \alpha)\Gamma(1 - \alpha)} \int_0^1 du u^{-\alpha}(1 - u)^{\alpha+\theta-1} \log u.$$

We do not know how to evaluate the integral using calculus but it is easy to evaluate numerically. For example of the use of Simpson's and Shannon's indices in studying breast cancer see [66]. Merlo et al [62] have investigated how well these indices and the Hill index

$$\left( \sum_i p_i^q \right)^{1/(1-q)}$$

perform as predictors of the progression of Barrett's esophagus to esophageal adenocarcinoma. They found that all of the diversity measures were strong and highly significant predictors of progression.

## References

- [1] Antal, T., and Krapivsky, P.L. (2011) Exact solution of a two-type branching process: models of tumor progression. *J. Stat. Mech.: Theory and Experiment* arXiv: 1105.1157
- [2] Armitage, P. (1952) The statistical theory of bacterial populations subject to mutations. *J. Royal Statistical Society, B.* 14, 1–40
- [3] Bailey, N.T.J. (1964) *The Elements of Stochastic Processes*. John Wiley and Sons, New York
- [4] Bozic I., Antal T., Ohtsuki H., Carter H., Kim D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci.* 107, 18545–18550
- [5] Crump. K.S., and Hoel, D.G. (1974) Mathematical models for estimating mutation rates in cell populations. *Biometrika.* 61, 237–252
- [6] Danesh, K., Durrett, R., Havriliesky, L., and Myers, E. (2013) A branching process model of ovarian cancer. *J. Theor. Biol.* 314, 10–15
- [7] Darling, D.A. (1952) The role of the maximum term in the sum of independent random variables. *Trans. American Math Society.* 72, 85–107



- [8] Durrett, R. (2008) *Probability Models for DNA Sequence Evolution*. Second Edition. Springer, New York
- [9] Durrett, R. (2010) *Probability: Theory and Examples*. Fourth edition. Cambridge U. Press
- [10] Durrett, R. (2012) Population genetics of neutral mutations in exponentially growing cancer cell populations. *Ann. Appl. Probab.* 23, 230–250
- [11] 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
- [12] Durrett, R., Foo, J., Leder, K., Mayberry, J., Michor, F. (2011) Intratumor heterogeneity in evolutionary models of tumor progression. *Genetics.* 188, 461–477
- [13] Durrett, R., and Moseley, S. (2010) Evolution of resistance and progression to disease during clonal expansion of cancer. *Theor. Popul. Biol.* 77, 42–48
- [14] Durrett, R., and Schweinsberg, J.. (2004) Approximating selective sweeps. *Theor. Popul. Biol.* 66, 129–138
- [15] Durrett, R., and Schweinsberg, J. (2005) Power laws for family sizes in a gene duplication model. *Ann. Probab.* 33, 2094–2126
- [16] Foo, Jasmine and Leder, Kevin (2013) Dynamics of cancer recurrence. *Ann. Appl. Probab.* 23, 1437–1468.
- [17] Foo, J., Leder, K., and Mummenthaler, S. (2013) Cancer as a moving target: understanding the composition and rebound growth kinetics of recurrent tumors. *Evolutionary Applications.* 6, 54–69
- [18] Fuchs, A., Joffe, A., and Teugels, J. (2001) Expectation of the ratio of the sums of squares to the square of the sum: exact and asymptotic results. *Theory Probab. Appl.* 46, 243–255
- [19] Griffiths, R.C., and Pakes, A.G. (1988) An infinite-alleles version of the simple branching process *Adv. Appl. Prob.* 20, 489–524
- [20] Haeno, H., Conen, M., Davis, M.B., Hrman, J.M., Iacobuzio-Donahue, C.A., and Michor, F. (2012) Computational modeling of pancreatic cancer reveals kinetics of metastasis suggesting optimum treatment strategies. *Cell.* 148, 362–375
- [21] Haeno, H., Iwasa, Y., and Michor, F. (2007) The evolution of two mutations during clonal expansion. *Genetics.* 177, 2209–2221

- [22] Haeno, H., and Michor, F. (2010) The evolution of tumor metastases during clonal expansion. *J Theor. Biol.* 263, 30–44
- [23] Harris, T.E. (1948) Branching processes. *Ann. Math. Statist.* 19, 474–494
- [24] Iwasa, Y., Nowak, M.A., and Michor, F. (2006) Evolution of resistance during clonal expansion. *Genetics.* 172, 2557–2566
- [25] Kesten, H. and Stigum, B. P. (1966) A limit theorem for multidimensional Galton-Watson processes. *Ann. Math. Statist.* 37, 1211–1223
- [26] Kingman, J.F.C. (1975) Random discrete distributions. *J. Royal Statistical Society, B.* 37, 1–22
- [27] Komarova, N.L., Wu, Lin, and Baldi, P. (2007) The fixed-size Luria-Delbruck model with a non-zero death rate. *Mathematical Biosciences.* 210, 253–290
- [28] Logan, B.F., Mallows, C.L., Rice, S.O., and Shepp, L.A. (1973) Limit distributions of self-normalized random sums. *Annals of Probability.* 1, 788–809
- [29] Lea, E.A., and Coulson, C.A. (1949) The distribution of the number of mutants in bacterial populations. *Journal of Genetics.* 49, 264–285
- [30] Leder, K., Foo, J., Skaggs, B., Gorre, M., Sawyers, C.L., and Michor, F. (2011) Fitness conferred by BCR-ABL kinase domain mutations determines the risk of pre-existing resistance in chronic myeloid leukemia. *PLoS One.* 6, paper e27682
- [31] Luria, S.E., and Delbruck, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics.* 28, 491–511
- [32] Michor, F, et al. (2005) Dynamics of chronic myeloid leukemia. *Nature.* 435, 1267–1270
- [33] O’Connell, N. (1993) Yule approximation for the skeleton of a branching process. *J. Appl. Prob.* 30, 725–729
- [34] Parzen, E. (1962) *Stochastic Processes.* Holden-Day, San Francisco
- [35] Pitman, J. (2006) *Combinatorial Stochastic Processes.* Springer, New York
- [36] Pitman, J., and Yor, M. (1997) The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of Probability.* 25, 855–900
- [37] Slatkin, M., and Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics.* 129, 555–562

- [38] Tomasetti, C., and Levy, D. (2010) Roles of symmetric and asymmetric division of stem cells in developing drug resistance. *Proc. natl. Acad. Sci.* 107, 16766–16771
- [39] Zheng, Q. (1999) Progress of a half-century in the study of the Luria-Delbrück distribution. *Mathematical Biosciences.* 162, 1–32
- [40] Zheng, Q. (2008) On Bartlett’s formulation of the Luria-Delbrück distribution. *Mathematical Biosciences.* 215, 48–54
- [41] Zheng, Q. (2009) Remarks on the asymptotics of the Luria-Delbrück and related distributions. *J. Appl. Prob.* 46, 1221–1224

### **Cancer Biology**

- [42] Armitage, P. (1985) Multistage models of carcinogenesis. *Environmental health Perspectives.* 63, 195–201
- [43] Armitage, P., and Doll, R. (1954) The age distribution of cancer and a multistage theory of carcinogenesis. *British J. Cancer.* 8, 1–12
- [44] Brown, P.O., and Palmer, C. (2009) The preclinical natural history of serous ovarian cancer: defining the target for early detection. *PLoS Medicine.* 6(7):e1000114.
- [45] Buys SS, Partridge E, Black A, et al. (2011) Effect of screening on ovarian cancer mortality The prostate, lung, colorectal and ovarian (PLCO) cancer screening randomized controlled trial. *JAMA* 305(22): 2295-2303. doi:10.1001/jama.2011.766.
- [46] Collisson, E.A., Cho, R.J., and Gray, J.W. (2012) What are we learning from the cancer genome? *Nature Reviews. Clinical Oncology.* 9, 621–630
- [47] Decruze, S.B., and Kirwan, J.M. (2006) Ovarian cancer. *Current Obstetrics and Gynecology.* 16(3): 161–167
- [48] Fearon, E.F. (2011) Molecular genetics of colon cancer. *Annu. Rev. Pathol. Mech. Dis.* 6, 479–507
- [49] Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell.* 87, 759–767
- [50] Feller, L., Kramer, B., and Lemmer, J. (2012) Pathobiology of cancer metastasis: a short account. *Cancer Cell International.* 12, paper 24
- [51] Fidler, I.J. (1978) Tumor heterogeneity and the biology of cancer invasion and metastases. *Cancer Research.* 38, 2651–2660

- [52] Fisher, J.C., and Holloman, J.H. (1951) A hypothesis for the origin of cancer foci. *British J. Cancer*. 7, 407–417
- [53] Fisher, R., Pusztai, L., and Swanton, C. (2013) Cancer heterogeneity: implications for targeted therapeutics. *Cancer Research*.
- [54] Gerlinger, M. et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England Journal of Medicine*. 366, 883–892
- [55] Knudson, A.G., Jr. (1971) Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci.* 68, 820–823
- [56] Knudson, A.G. (2001) Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*. 1, 157–162
- [57] Jemal, A., Siegel, R., Xu, J., and Ward, E. (2010) Cancer statistics, 2010. *CA: A Cancer Journal for Clinicians* 60(5): 277–300
- [58] Jones, S., et al. (2008) Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science*. 321, 1801–1812
- [59] Lengyel, E. (2010) Ovarian cancer development and metastasis. *The American Journal of Pathology*. 177(3): 1053–1064.
- [60] Luebeck, E.G., and Mollgavkar, S.H. (2002) Multistage carcinogenesis and the incidence of colorectal cancer. *proc. natl. Acad. Sci.* 99, 15095–15100
- [61] Maley, C.C., et al. (2006) Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nature Genetics*. 38, 468–473
- [62] Merlo, L.M.F., et al (2010) A comprehensive survey of clonal diversity measures in Barrett’s esophagus as biomarkers of progression to esophageal adenocarcinoma. *Cancer Prevention Research*. 3, 1388–
- [63] Naora, H., and Montell, D.J. (2005) Ovarian cancer metastasis: integrating insights from disparate model organisms. *Nature Reviews Cancer*. 5(5): 355–366
- [64] Navin, N., et al (2011) Tumor evolution inferred from single cell sequencing. *Nature*. 472, 90–94
- [65] Nordling, C.O. (1953) A new theory on cancer inducing mechanism. *British J. Cancer*. 7, 68–72
- [66] Park, S.Y., Gönen, M, Kim, H.J., Michor, F., and Polyak, K. (2010) Cellular and genetic diversity in the progression of in situ human breast cancer to an invasive phenotype.

- [67] Parsons, D.W., et al. (2008) An integrated genomic analysis of human glioblastoma multiforme. *Science*. 321, 1807–1812
- [68] Russnes, H.G., Navin, N., Hicks, J., and Borrensen-Dale, A.L. (2011) Insight into the heterogeneity of breast cancer inferred through next generation sequencing. *J. Clin. Invest.* 121, 3810–3818
- [69] Schorge, J.O., Modesitt, S.C., Coleman, R.L., Cohn, D.E., Kauff, N.D., Duska, L.R., and Herzog, T.J. (2010) SGO White Paper on ovarian cancer: etiology, screening and surveillance. *Gynecologic Oncology*. 119(1): 7–17.
- [70] Siegel, R., Naishadham, D., and Jemal, A. (2012) Cancer statistics, 2012. *CA: A Cancer Journal for Clinicians*. 62: 1029. doi: 10.3322/caac.20138
- [71] Sjöblom, T., et al. (2006) The consensus coding sequences of human breast and colorectal cancers. *Science*. 314, 268–274
- [72] Sottoriva, A., et al. (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* 110, 4009–4014
- [73] Surveillance, Epidemiology, and End Results (SEER) Program. <http://seer.cancer.gov/>.
- [74] The Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 455, 1061–1068
- [75] Tomasetti C., Vogelstein, B., and Parmigiani, G. (2013) Half or more somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proc. Natl. Acad. Sci.* 110, 1999–2004
- [76] Valastyan, S., and Weinberg, R.A. (2011) Tumor metastasis: Molecular insights and evolving paradigms. *Cell*. 147, 275–292
- [77] Wood, L.D., et al. (2007) The genomic landscapes of human breast and colorectal cancers. *Science*. 318, 1108–1113