

# Branching processes, budding yeast, and $k$ -nacci numbers

Peter Olofsson\*

March 26, 2010

## Abstract

A real-world application of branching processes to a problem in cell biology where generalized  $k$ -nacci numbers play a crucial part is described. The  $k$ -nacci sequence is used to obtain computational formulas and to justify certain practical simplifications.

*Keywords:* Branching process, yeast,  $k$ -nacci sequence *AMS 2000 Mathematics Subject Classification Codes:* 11B39, 60J80, 62P10, 92D25

## 1 Introduction

The Fibonacci sequence is famous for showing up in nature in a multitude of ways, some of which are however idealized such as Fibonacci's breeding rabbits. In Olofsson and Bertuch (2010), a branching process model was used to analyze experiments on growing yeast populations and it turned out that generalized  $k$ -nacci numbers were crucial to practical results and calculations. In this paper, we elaborate on some of the mathematics that were not presented in detail in Olofsson and Bertuch, the focus of that paper being on solving a problem in biology.

By the  $k$ -nacci sequence we shall in this paper mean the sequence  $\{F_j, j \geq 0\}$  defined by  $F_0 = F_1 = \dots = F_{k-1} = 0$ ,  $F_k = 1$ , and  $F_n = F_{n-1} + F_{n-2} +$

---

\*Trinity University, Mathematics Department, One Trinity Place, San Antonio, TX 78212. Email: polofso@trinity.edu

$\dots + F_{n-k}$  for  $n > k$ . We will utilize results from Flores (1967) regarding the asymptotics of the sequence. Before we get to the main results, let us present a brief overview of branching processes, followed by a quick look at its application to yeast populations.

A branching process is a stochastic model for a proliferating population where assumptions are made on individual reproduction to draw conclusions about population behavior. One of the simplest cases is to consider a population of cells or bacteria that reproduce by binary splitting after going through the cell cycle. Thus, an individual always gives rise to two daughters and randomness enters via the cell cycle times (or “lifetimes”) which are assumed to be independent random variables with common cumulative distribution function (cdf)  $F$ . Start such a population from one individual and let  $M(t)$  denote the expected number of individuals at time  $t$ .

To arrive at an expression for  $M(t)$ , first note that there are  $2^n$  cells in the  $n$ th generation. Such a cell is present at time  $t$  if the sum of  $n$  cell cycle times is less than  $t$  and the sum of  $n + 1$  cell cycle times is greater than  $t$ . The sum of  $j$  cell cycle times has a cdf that is the  $j$ -fold convolution of  $F$  with itself, denoted  $F^{*j}$ . Thus, the probability that an  $n$ th-generation cell is present at time  $t$  equals  $F^{*n}(t) - F^{*(n+1)}(t)$  and we get the expression

$$M(t) = 1 - F(t) + \sum_{n=1}^{\infty} 2^n \left( F^{*n}(t) - F^{*(n+1)}(t) \right) \quad (1.1)$$

where the first term  $1 - F(t)$  accounts for the original cell.

In Olofsson and Bertuch (2010) a population of the budding yeast *Saccharomyces cerevisiae* was studied, focusing on describing growth rates in response to the attrition and maintenance of telomeres (chromosomal ends). For yeast, the situation is more complicated than above since a yeast cell reproduces not by splitting but by budding. Thus, there is a clear distinction between mother and daughter, and it is known that a mother cannot give birth to an unlimited number of daughters. Thus, rather than  $2^n$  cells in the  $n$ th generation, we get a number  $m(n)$  and the expression

$$M(t) = 1 - F(t) + \sum_{n=1}^{\infty} m(n) \left( F^{*n}(t) - F^{*(n+1)}(t) \right) \quad (1.2)$$

Growth rate and other asymptotic properties of  $M(t)$  are closely related to those of  $m(n)$ . Also, for applications, we need to compute  $M(t)$  for finite  $t$

which requires getting a handle on the number  $m(n)$ . We shall devote the rest of this paper to this task.

Denote the number of daughter cells by  $k$ , termed the proliferative lifespan. In any given generation  $n$ , cells can be divided into classes describing how many more daughter cells they can have. Thus, let  $n_j$  be the number of cells in generation  $n$  that can have an additional  $j$  daughter cells for  $j = 0, 1, \dots, k$ . The class with  $j = 0$  are the non-proliferating cells and we assume that they stay in the population indefinitely (although it is easy to model a scenario where they eventually die and disappear). The class with  $j = k$  are the newborn cells that have yet to reproduce.

As it turns out, the numbers of cells in these classes are precisely described by the  $k$ -nacci sequence.

**Proposition 1.1** *Consider the vector  $(n_0, n_1, \dots, n_k)$  in the  $n$ th generation of the branching process above. Let  $F_i$  denote the  $i$ th  $k$ -nacci number and let  $S_n = F_0 + F_1 + \dots + F_n$ . Then  $(n_0, n_1, \dots, n_k)$  equals*

$$(S_n, F_{n+1}, F_{n+2}, \dots, F_{n+k})$$

*Proof.* Let  $n_i$  be the number of cells that are able to reproduce another  $i$  times for  $i = 0, 1, \dots, k$ . Each cell with  $i > 1$  produces a daughter cell that is able to reproduce  $k$  times and is then itself able to reproduce another  $i - 1$  times. Cells with  $i = 0$  remain unchanged. In generation 0 there is one cell that is able to divide  $k$  more times which gives the vector  $(0, 0, \dots, 0, 1)$  for generation 0. The transition from generation  $n - 1$  to generation  $n$  is as follows:

$$\left\{ \begin{array}{l} \text{Generation } n - 1 : (n_0, n_1, \dots, n_{k-1}, n_k) \\ \text{Generation } n : (n_0 + n_1, n_1 + n_2, \dots, n_{k-1} + n_k, \sum_{i=1}^k n_i) \end{array} \right.$$

and the proposition follows. ■

Proposition 1.1 provides a recursive scheme that enables us to compute the number of cells  $n_0 + n_1 + \dots + n_k$  in each generation and by (1.2), the expected number of cells at each time  $t$ . To establish the long-term composition of cells in the different classes, we rely on the fact that  $k$ -nacci numbers have

asymptotic geometric growth. Following Flores (1967), there exists a number  $r$  such that

$$F_j \sim r^{j-k}$$

for  $j$  large enough. The number  $r$  is the largest root to the equation

$$x^k - x^{k-1} - \dots - x - 1 = 0$$

which is known to be real and lie between the golden ratio  $\phi$  and 2. In fact, for  $k = 2$ ,  $r = \phi$  and  $r \uparrow 2$  as  $k \rightarrow \infty$ . The asymptotic proportions are given next.

**Proposition 1.2** *Let  $r$  be as above. The asymptotic proportions of cells in the classes  $(0, 1, \dots, k)$  equal*

$$(r^{-k}, (r-1)r^{-k}, (r-1)r^{-(k-1)}, \dots, (r-1)r^{-2}, (r-1)r^{-1})$$

*Proof.* By Proposition 1.1, the proportions equal

$$\left( \frac{S_n}{S_{n+k}}, \frac{F_{n+1}}{S_{n+k}}, \dots, \frac{F_{n+k}}{S_{n+k}} \right)$$

and by Flores (1967), we have the asymptotic behavior  $F_j \sim r^{j-k}$ . Hence,

$$S_n = \sum_{j=0}^n F_j \sim \frac{r^{-k}}{k-1} r^{n+1}$$

which also gives

$$S_{n+k} \sim \frac{r^{-k}}{k-1} r^{n+k+1} = \frac{r^{n+1}}{k-1}$$

and hence

$$\frac{S_n}{S_{n+k}} \sim r^{-k}$$

and

$$\frac{F_{n+j}}{S_{n+k}} \sim (r-1)r^{j-(k+1)}$$

for  $j = 1, \dots, k$  which proves the proposition. ■

As a corollary we also obtain the growth rate of  $M(t)$ . To that end, let  $\widehat{F}$  denote Laplace transform, that is,

$$\widehat{F}(s) = \int_0^\infty e^{-st} F(dt)$$

to obtain the following result.

**Corollary 1.3** *As  $t \rightarrow \infty$ ,  $M(t) \sim Ce^{\alpha t}$  where  $\alpha > 0$  is the solution to the equation  $r\widehat{F}(\alpha) = 1$  and  $C$  is a constant that depends on  $r, k$ , and the cdf  $F$ .*

*Proof.* From the proof of Proposition 1.2, we get

$$m(n) = S_{n+k} \sim \frac{r^{n+1}}{k-1}$$

If we substitute this expression for  $m(n)$  in (1.2), we get

$$M(t) \sim \frac{r}{k-1} \sum_n r^n \left( F^{*n}(t) - F^{*(n+1)}(t) \right)$$

for large  $t$ . The asymptotics of the sum are obtained by standard branching process methods for a binary splitting process with mean number of offspring equal to  $r$  and can be shown to equal  $be^{\alpha t}$  where  $\alpha > 0$  is the solution to the equation  $r\widehat{F}(\alpha) = 1$  (note that  $r > 1$  which implies that  $\alpha > 0$ ). For details, see Harris (1963) or Jagers and Nerman (1984). The constant  $b$  equals

$$b = \left( 4\alpha \int_0^\infty te^{-\alpha t} F(dt) \right)^{-1}$$

and hence  $C = br/(k-1)$ . ■

Proposition 1.2 is not just a theoretical limit result, it has important practical implications for the yeast cell population studies. Certain computational expressions become greatly simplified if the finite proliferative lifespan can be neglected, instead assuming that each cell can produce an unlimited number of daughter cells. Since the fraction of non-proliferating cells in a given generation  $n$  is roughly  $r^{-k}$ , this number can be used to justify such an approximation. For example, for the regular Fibonacci sequence with  $k = 2$ , we have  $r = \phi$  and since  $\phi^{-2} \approx 0.38$ , as many as 38% of cells have reached the

end of their proliferative lifespan and are no longer able to produce daughter cells. In this case, the approximation would not work very well. Note that, since  $r \uparrow 2$  as  $k \rightarrow \infty$ , the “ $\infty$ -nacci” sequence has  $r = 2$  and corresponds to a binary splitting branching process where each individual can produce an unlimited number of offspring.

For yeast cells, the proliferative lifespan  $k$  has been estimated to be on average 25, Sinclair *et al.* (1998), which gives a value of  $k$  that for all practical purposes equals 2 and the fraction of non-proliferating cells is less than one in 10 million. For any reasonable duration of a yeast cell experiment, this fraction is negligible although it does of course matter to the theoretical asymptotic limits. The number 25 is an experimentally determined average and the true range may well go lower. However, calculations show that  $r$  exceeds 1.99 already for  $k = 7$  in which case less than 1% of cells are non-proliferating. For the yeast experiments considered in Olofsson and Bertuch (2010),  $k$  is likely to largely exceed 7 and the approximation works well.

## 2 References

- [1] I. Flores, *Direct calculation of  $k$ -generalized Fibonacci numbers*. The Fibonacci Quarterly **5** (1967), 259–266
- [2] T.E. Harris, *The Theory of Branching Processes*. Springer, Berlin, 1963.
- P. Jagers and O. Nerman, *The Growth and Composition of Branching Populations*. Adv. Appl. Prob. **16** (1984), 221–259
- P. Olofsson and A.A. Bertuch, *Modeling growth and telomere dynamics in *S. cerevisiae**. J. Theor. Biol. **263(3)** (2010), 353–359
- D.A. Sinclair, K. Mills and L. Guarente, *Aging in *Saccharomyces cerevisiae**. Annu. Rev. Microbiol. **52** (1998), 533–560