

# Branching Processes and Non-Commuting Random Variables in Population Biology

Timothy C. Reluga

August 4, 2010

## Abstract

Branching processes are a well-established tool in mathematical biology used to study the dynamics of rarefied populations where agents act independently and small stochastic density-independent changes in population sizes. However, they are often avoided by non-mathematicians because of their reliance on generating functions. Generating functions are powerful computational aids but are often difficult to motivate. In this paper, I review branching process theory using a non-commuting random variable description of multiplication as mnemonic for generating functions. Starting from the elementary definition of multiplication, I show how uncertainty leads to a natural generalization of integer multiplication without the commutative property, and how this in-turn is connected to the well-established study of generating functions. Non-commuting random-variable methods are described in detail and illustrated using examples.

## 1 Introduction

Branching process theory is one of the oldest [Galton, 1889] and richest [Ulam, 1990, Harris, 1963, Mode, 1971, Athreya and Ney, 1972, Kimmel and Axelrod, 2002, Macken and Perelson, 1985, Haccou et al., 2005, Dorman et al., 2004] topics in mathematical population biology. A branching process is a stochastic process in which all individuals reproduce randomly, independent of interactions with other individuals. Originally motivated by problems concerning the extinction of surnames in Europe, branching process theory now plays a role in contemporary biology fields as diverse as carcinogenesis [Meza et al., 2008], network theory [Newman et al., 2001],

evolution [Reluga et al., 2007], and botany [Loi and Cournède, 2008, Prusinkiewicz et al., 2007].

Despite its power and pedigree, branching process methods are not as widely employed by researchers as they could be. One reason for the lack of employment may be that the domains of application for branching process methods are smaller than the domains for other methods like calculus, matrix algebra, or graph theory. Another possible reason for the neglect of branching processes is that their mathematical properties are not simple enough to be covered in the core college mathematics curriculum.

The study of branching processes usually revolves around mathematical objects called probability generating functions that represent the probabilities or expectations of different events. Probability generating functions can make certain complex calculations trivial. However, they are usually formulated in terms of “dummy” variables that have no obvious motivations, and are more naturally investigated in terms of moment statistics rather than common cumulant statistics like variance and skew.

Such pedagogical barriers to the adoption of computational tools are not unknown. In *Principia Mathematica* [Newton, 1686], for instance, Newton presented geometric constructions for many results he is believed to have obtained using calculus. Today, few students can repeat Newton’s geometric calculations, but many can get those results using calculus. Similarly, actuarial calculations were commonly conducted using life-tables before Alfred Lotka introduced the use of matrix algebra [Caswell, 2001]. Today, matrix algebra is common in biology. It’s applications include the growth of stage-structured populations, covariance matrices, and Markov-process descriptions of genetic mutation. Branching processes are not as powerful as either calculus or matrices, but are useful for some problems. An alternative pedagogical approach to branching processes may facilitate their wider use.

In this paper, I will present a review of branching process theory that highlights the algebraic structure of branching processes in terms of non-commuting random variables ( NCRVs ), and shows how this algebraic structure can be interpreted directly in terms of both modelling concepts and cumulant generating functions. The paper is intended for mathematical-biology graduate students seeking an alternative introduction to branching processes. Almost all of the mathematical results are classical and have been published elsewhere. The original contribution of this paper is the interpretation of branching processes in terms of multiplication of random variables. First, we investigate the concept of multiplication under uncertainty. The properties of multiplication under uncertainty are then shown to naturally coincide with the rules of cumulant generating function algebra. This coincidence is used to derive some standard results.

## 2 Non-commuting random variables

After learning addition, elementary school students in the United States learn how to multiply numbers. Multiplication, as we first learn, is repeated addition. Suppose we have a tree with  $X = 3$  tree branches and each branch has  $Y = 5$  apples. How many apples do we have in total? Well,

$$X \times Y = 3 \text{ branches} \times 5 \text{ apples per branch} = 5 + 5 + 5 = 15 \text{ apples.} \quad (1)$$

In the same way,

$$4 \text{ branches} \times 5 \text{ apples per branch} = 5 + 5 + 5 + 5 = 20 \text{ apples.} \quad (2)$$

Repeated addition becomes tedious, however, as numbers become large. Soon, students begin to memorize their multiplication tables for numbers 1 to 10. At this point, students start to forget about order. It doesn't matter if we have 3 branches with 5 apples each or 5 branches with 3 apples each. Either way,  $3 \times 5 = 5 \times 3 = 15$  apples in total. The clever student may point out that this isn't quite true, since we have a different number of branches in the two cases, but such mild protests seldom can stand against the lock-step of common curricula. As students move on to more advanced concepts like fractions, decimal numbers, and complex numbers, the fundamental definition given in Eq. (1) fades into the background.

Now, suppose that (as is almost always the case) there is some uncertainty about the number of branches, and the number of apples on each branch. For instance, suppose that there is an 80 percent chance of the tree having 3 branches, but a 10 percent chance that there are 2 branches and a 10 percent change that there are 4 branches. And suppose that each branch has equal probability of having 4, 5, or 6 apples. How many apples in total are there?

Obviously, this question cannot be completely answered with a single number. The number of branches and the number of apples per branch are uncertain; they are "random variables". We might guess that there should be about 15 apples in total, but there might be as few as  $2 \times 4 = 8$  apples or as many as  $4 \times 6 = 24$  apples.

Experimentally inclined readers might throw some dice to gain some intuition.

1. Obtain 1 ten-sided die with sides numbered 1 to 10, and 4 3-sided dice with sides numbered 4,5, and 6.
2. Roll a 10-sided die.
3. If the ten-sided die shows 2 or 4, roll 2 or 4 3-sided dice respectively. Otherwise, roll 3 3-sided dice.

4. Add up all the 3-sided dice to get a guess at the total number of apples.

Repeating this sequence numerous times physically or in pseudorandom-number simulations, we can get a histogram for the likelihoods of various outcomes.

There is a mathematical calculation that can precisely express our uncertainty about the total number of apples if we know the uncertainty in the number of branches and the number of apples per branch. For the moment, we will postpone explanation of the calculation, but its result is shown in Figure 1A. The most likely number of apples is 15, and all of the weight of the distribution is clustered tightly around the maximum. If there are not 15 apples, our next best guesses would be 14 or 16 apples.

But what if the numbers and their uncertainty are reversed? Suppose a branch contains 3 apples 80 percent of the time, 2 apples 10 percent of the time, and 4 apples 10 percent of the time, while there is a one-third chance of having 4, 5, or 6 branches. In this case, we still expect there to be about 15 apples in total with no more than 24 and no fewer than 8, but our uncertainty is distributed in a different manner than before. Our uncertainty is now concentrated in the number of branches instead of the number of apples per branch. 12 and 18 are our next-to-best guesses now. While the graph of the probabilities in the preceding example had a single peak (Figure 1A), this example has 3 strongly-separated peaks (Figure 1B). Thus, reversing the order of multiplication has significantly altered the nature of our uncertainty.

These heuristics can easily be turned into mathematics. To begin with, we will restrict our number system to the counting numbers  $\mathbb{N} = 0, 1, 2, \dots$ . A non-commuting random variable (NCRV)  $x$  represents a counting number whose value is uncertain<sup>1</sup>. Associated with the NCRV  $x$  is a probability distribution  $p_x(n)$  that describes the probability that any given realization of the random variable has value  $n$ . The algorithmically inclined can think of  $x$  as a call to a random number generator in a computer program that returns independent samples from the counting number distribution  $p_x$ . If all the weight of the distribution  $p_x$  is on a single value, then  $x$  is a deterministic variable that returns the same value for all realizations. Since each realization of the variable  $x$  may give a different value, it is not always true that  $x$  is equal to itself. But we can define a congruence equivalence relation where  $x \sim y$  if and only if for all  $n$ ,  $p_x(n) = p_y(n)$  for every  $n$  such that  $x \sim x$  despite sometimes having  $x \neq x$ .

NCRVs can be added, of course. Suppose  $x$ ,  $y$ , and  $z$  are NCRVs. (Through-out this paper, all NCRVs will be assumed to be mutually independent.) Then  $x + y \sim z$

---

<sup>1</sup>The reason for this name will become apparent soon.

if and only if for every  $n$ ,

$$p_z(n) = \sum_{i=0}^n \text{Prob}(x = i \text{ and } y = n - i) = \sum_{i=0}^n p_x(i)p_y(n - i). \quad (3)$$

More interestingly, NCRVs can also be multiplied in the sense that

$$xy = \begin{cases} 0 & \text{if } x = 0, \\ \sum_{i=1}^x y_i & \text{if } x > 0 \end{cases} \quad (4)$$

where each  $y_i$  is an independent realization of  $y$ . Then it is a matter of algebra to show that  $z \sim xy$  if and only if for all  $n$ ,

$$p_z(n) = \sum_{i=0}^{\infty} \left[ p_x(i) \sum_{n_1+n_2+\dots+n_i=n} \left( \prod_{j=1}^i p_y(n_j) \right) \right] \quad (5)$$

where  $n_j \geq 0$  for all  $j$ . This is the formula used to draw Figure 1. Considering the asymmetry in the definition of multiplication, it is not surprising that multiplication of NCRVs with uncertainty is not commutative and  $xy \approx yx$  in general. A related consequence is that

$$(x + y)z \sim xz + yz, \quad (6)$$

but in general

$$z(x + y) \approx zx + zy. \quad (7)$$

Thus, NCRVs are right-distributive, but not left-distributive. In (6), both sides lead to one realization of  $x$  and  $y$  and the same numbers of realizations of  $z$ . In (7),  $x$  and  $y$  always have the equal numbers of realizations on the left but may have unequal numbers of realizations on the right.

While the definitions of addition and multiplication are similar to those in common usage, we do not have corresponding division and subtraction operations. Unless an NCRV  $x$  is deterministic, there are no additive inverses  $-x$  such that  $(-x) + x = 0$  and no multiplicative inverses  $x^{-1}$  such that  $xx^{-1} = 1 = x^{-1}x$ . Uncertainty makes it impossible to precisely invert an NCRV.

Thus, we find that the clever student is more insightful than we or they realized. The property that  $F \times G = G \times F$  is called the commutative property of multiplication. Commutativity is not a universal property of multiplication, but one

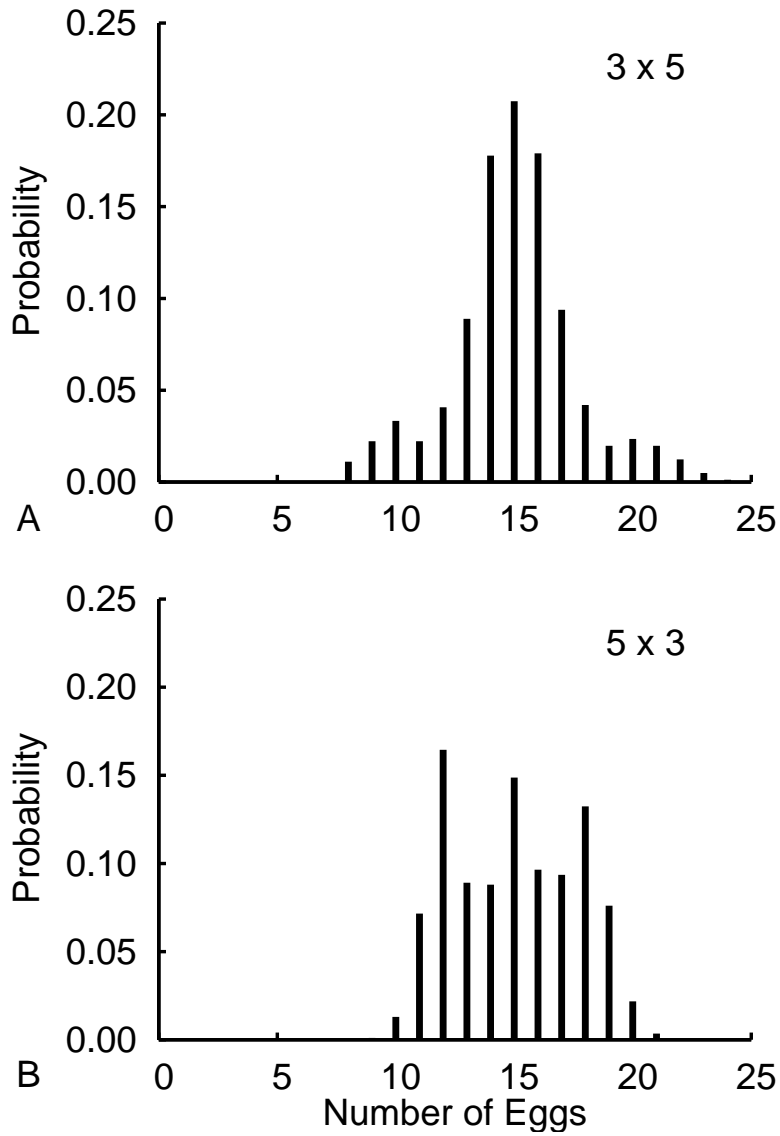


Figure 1: Probability distribution for the total numbers of apples when we are uncertain. The probability distributions of the products “ $3 \times 5$ ” (A) and “ $5 \times 3$ ” (B) with uncertainty as described in the text. Both have the same expected value of 15, but strikingly different shapes. The distribution of  $3 \times 5$  has a unique peak, but the distribution of  $5 \times 3$  has 3 peaks. The distribution of  $5 \times 3$  is more semi-circular than that of  $3 \times 5$ .

that holds only for certain number systems like the integers and the real numbers. Students seldom encounter non-commutative systems until they are introduced to matrices in college. But noncommutative number systems are not actually exotic, and uncertainty is one reason they appear in practice.

The initial example of tree branches and apples implicitly considers two different types of objects. Some of the conceptual questions regarding the meaning of reversing the order of multiplication might be resolved if the types being multiplied were made explicit. As it turns out, we can extend our concept of multiplication to cases where individuals/particles have multiple types.

In order to preserve parallels to matrix theory, we will consider only random variables whose output types are the same as their input types. Suppose we have a process where individuals can have types  $1 \dots m$ . Define a multitype NCRV  $F$  as an  $m \times m$  matrix where the row vector  $F_i$  is a random variable returning a number of individuals of each type  $j = 1 \dots m$  generated from one individual of type  $i$ . The entries of  $F_i$  may be correlated, so  $F_{ij}$  and  $F_{ik}$  are often dependent for any  $i, j, k$ . As before, rows are congruent to themselves ( $F_i \sim F_i$ ) although realizations of a row may not be equal ( $F_i \neq F_i$ ). Different rows of  $F$  are always independent, so  $F_{ij}$  and  $F_{kl}$  are independent provided  $i \neq k$ .

Let  $G$  be a different multitype NCRV on  $1 \dots m$ . Then each row  $i$  of the product  $FG$  is given by

$$(FG)_i := \sum_{j=1}^m \sum_{k=1}^{F_{ij}} G_{j,k} \tag{8}$$

where for each  $k$ ,  $G_{j,k}$  is an independent realization of the random vector  $G_j$ . The computational implementation to calculate a realization of the random matrix  $FG$  is:

1. For each  $i$ , calculate a realization of the random vector  $F_i$ .
2. For each  $j$ , calculate  $F_{ij}$  realizations of the vector random variable  $G_j$  and add them all together (In the degenerate case of  $F_{ij} = 0$ , the sum is the 0-vector.)

This can easily be extended to cases where the initial types differ from the output types. All finite-state discrete-time Markov processes can be included as a special case of multitype multiplication. The same algebraic properties that we've mentioned for scalar NCRVs apply to multitype NCRVs. With a little care, we can even extend these ideas to continuously-typed processes where the types are locations in spaces with real-number coordinates.

The reader should note that the convention we have adopted interpreting NCRVs as having independent rows leads to a natural left-to-right interpretation of matrix multiplication. This is particularly convenient as it will correspond naturally with the standard function notation we will use below for generating functions. While this is the standard convention in probability-literature, it is the opposite of the convention in applied mathematics, where column bases are used and matrix equations are most often read right-to-left.

### 3 Generating Functions

The algebraic formulation of a language of NCRVs is pedagogically appealing, but also needs to be associated with a methodology for calculation that leads from expressions to results. A convenient and powerful computational methodology, referred to by one author as “generatingfunctionology” [Wilf, 1999], the study of generating functions, fills this need. Generatingfunctionology makes use of a special class of non-linear functions, called generating functions, to simplify certain common calculations in a manner similar to the use of matrices to simplify the study of linear functions. Probability theory has made use of generating functions since its inception, beginning with the work of De Moivre in 1730 and making a significant appearance in Laplace’s derivation of the Central Limit Theorem in 1810 [Stigler, 1986]. Generating functions are powerful because they implicitly encapsulate the basic sorting and searching operations needed to simplify mathematical expressions. This has made their use popular in many areas of mathematics. Generating functions are convenient because they allow us to use standard calculus operations to determine common results like the expectation of a random variable. This section defines generating functions and reviews some of their their basic properties as they relate to NCRVs . In particular, we highlight the relationship between NCRVs and a particular flavor of generating functions call cumulant generating functions.

A generation function is a representation of the probability distribution of a random variable, put in a convenient mathematical form [Kimmel and Axelrod, 2002, Haccou et al., 2005]. There are several different flavors of generating functions. We will begin with one called “cumulant generating functions” (CGFs). If  $x$  is our random variable, the cumulant generating function of  $x$  is defined as

$$\tilde{x}(\theta) := \ln\langle e^{x\theta} \rangle \tag{9}$$

where  $\langle \cdot \rangle$  represents the expected value of its argument. If the random variable is a

non-negative integer, then

$$\tilde{x}(\theta) = \ln \sum_{n=0}^{\infty} p_x(n) e^{n\theta}, \quad (10)$$

Here, the natural log should be interpreted in a symbolic sense as the inverse of function of the exponential such that

$$e^{\tilde{x}(\theta)} = \sum_{n=0}^{\infty} p_x(n) e^{n\theta} \quad (11)$$

and the CGF is uniquely defined.

Cumulant generating functions get their names from their property of generating the mean, variance, and higher order cumulant statistics of their underlying distribution. Specifically, the coefficients of the cumulant generating function's Taylor series around  $\theta = 0$  are given by

$$\tilde{x}(\theta) = \langle x \rangle \theta + \frac{1}{2} \langle \langle x \rangle \rangle \theta^2 + O(\theta^3), \quad (12)$$

where  $\langle \langle \cdot \rangle \rangle$  represents the variance. The expected value of  $x$  is

$$\tilde{x}'(\theta = 0) = \frac{\sum_{n=0}^{\infty} n p_x(n)}{\sum_{n=0}^{\infty} p_x(n)} = \langle x \rangle, \quad (13)$$

the variance of  $x$  is  $\tilde{x}''(\theta = 0)$ , and higher order cumulants can be calculated in a similar manner. Note that as long as the probabilities  $p_x(n)$  sum to 1,  $\tilde{x}(0) = 0$ . If a variable  $k$  is deterministic, such that

$$p_k(n) = \begin{cases} 1 & \text{if } n = \hat{k}, \\ 0 & \text{otherwise,} \end{cases} \quad (14)$$

then its CGF

$$\tilde{k}(\theta) = \ln \sum_{n=0}^{\infty} p_k(n) e^{n\theta} = \ln e^{\hat{k}\theta} = \hat{k}\theta. \quad (15)$$

All cumulants of a deterministic variable are zero except for the expectation. This matches our intuition, since we do not think of the number 3 as having a variance or skew.

CGFs are just one of several closely related families of generating functions including the moment generating function (MGF)  $\langle e^{\theta x} \rangle$ , the Z-transform  $\langle \lambda^{-x} \rangle$ , the characteristic function and Fourier transform  $\langle e^{itx} \rangle$ , and the probability generating function (PGF)  $\langle s^x \rangle$ . Each of these has a natural application. Using the PGF, for instance, the probability  $p_n$  of any integer  $n$  can be calculated by interpreting the argument as a complex number and evaluating a contour integral,

$$p_n = \frac{1}{(2\pi i)^{n+1}} \oint_{s=0} \frac{\langle s^n \rangle}{s^{n+1}} ds. \quad (16)$$

All four of these flavors of generating functions are formally interrelated by various transforms, so we can choose one as a standard for our analysis. CGF's are special for one particular reason: the algebraic rules of cumulant generating functions exactly correspond to the algebraic rules of the random variables they describe. We can show this by proving a correspondence between the various common algebraic operations. Take  $w, x, y$ , and  $z$  as NCRVs with CGFs  $\tilde{w}(\theta) = \ln \langle e^{w\theta} \rangle$ ,  $\tilde{x}(\theta) = \ln \langle e^{x\theta} \rangle$ ,  $\tilde{y}(\theta) = \ln \langle e^{y\theta} \rangle$ , and  $\tilde{z}(\theta) = \ln \langle e^{z\theta} \rangle$ .

**Proposition 1.** *If  $z \sim x + y$ , then  $\tilde{z}(\theta) = \tilde{x}(\theta) + \tilde{y}(\theta)$ .*

*Proof.* By definition,

$$\tilde{z}(\theta) = \ln \langle e^{z\theta} \rangle = \ln \langle e^{(x+y)\theta} \rangle = \ln \langle e^{x\theta} e^{y\theta} \rangle. \quad (17a)$$

Since  $x$  and  $y$  are independent variables,

$$\ln \langle e^{x\theta} e^{y\theta} \rangle = \ln [\langle e^{x\theta} \rangle \langle e^{y\theta} \rangle] = \ln \langle e^{x\theta} \rangle + \ln \langle e^{y\theta} \rangle. \quad (17b)$$

We now conclude that

$$\tilde{z}(\theta) = \tilde{x}(\theta) + \tilde{y}(\theta). \quad (17c)$$

□

NCRVs and their CGF's also share a multiplicative relationship.

**Proposition 2.**  *$z \sim xy$  implies  $\tilde{z}(\theta) = \tilde{x}(\tilde{y}(\theta))$ .*

*Proof.* We will need to make use of two particular results in this proof. The first is a special case of our general result. If  $x$  is actually a constant in the sense that it

takes on a particular integer value with probability 1, then since all realizations of  $y_i$  are independent,

$$\langle e^{(\sum_{i=1}^x y_i)\theta} \rangle = \langle e^{y_1\theta} e^{y_2\theta} \dots e^{y_x\theta} \rangle = \langle e^{y_1\theta} \rangle \langle e^{y_2\theta} \rangle \dots \langle e^{y_x\theta} \rangle = \langle e^{y\theta} \rangle^x. \quad (18a)$$

Second is the idea of conditional expectation [Williams, 1991]. In general, the expectation of a function of a set of random variables can be rewritten as the expectation over the first of the conditional expectation over the rest given the first. Specifically,

$$\langle f(x, y_1, \dots) \rangle = \langle \langle f(x, y_1, \dots) | x \rangle \rangle. \quad (18b)$$

We now proceed to the result. By definition,

$$\tilde{z}(\theta) = \ln \langle e^{z\theta} \rangle = \ln \langle e^{(\sum_{i=1}^x y_i)\theta} \rangle = \ln \langle \langle e^{(\sum_{i=1}^x y_i)\theta} | x \rangle \rangle. \quad (18c)$$

By Eq. (18a)

$$\langle e^{(\sum_{i=1}^x y_i)\theta} | x \rangle = \langle e^{y\theta} \rangle^x, \quad (18d)$$

so

$$\tilde{z}(\theta) = \ln \langle \langle e^{y\theta} \rangle^x \rangle = \ln \langle e^{x \ln \langle e^{y\theta} \rangle} \rangle = \tilde{x}(\tilde{y}(\theta)). \quad (18e)$$

□

So multiplication of NCRVs corresponds to composition of their CGF's. That the converses are also true follows from the uniqueness of CGF's.

**Proposition 3.** *If  $x$  and  $z$  are random variables on the counting numbers and  $\tilde{z}(\theta) = \tilde{x}(\theta)$ , then  $z \sim x$ .*

*Proof.* Restricting  $\theta$  to the positive real numbers, there is a bijection for CGFs to PGFs. Given PGFs, we can then show  $p_z(n) = p_x(n)$  for all  $n$ , implying  $z \sim x$ . □

Clearly in these proofs, I am sacrificing rigor for simplicity. Generating functions are primarily symbolic objects or data-structures to borrow a term from computer-science. Despite the obvious temptations, interpretations of generating functions as “functions” on the real or complex numbers are secondary.

## 4 Elementary Applications

The correspondence between the random variables and their CGFs allows us to quickly derive certain results. An elementary algebraic result is that NCRV multiplication is associative.

**Proposition 4.** *If  $x$ ,  $y$ , and  $z$  are NCRVs,  $x(yz) \sim (xy)z$ .*

*Proof.* Let  $u \sim x(yz)$ . Then  $\tilde{u}(\theta) = \tilde{x}(\tilde{y}(\tilde{z}(\theta)))$ . Since function composition is always associative, we can either calculate the composition of  $\tilde{y}$  and  $\tilde{z}$ , and then compose with  $\tilde{x}$  or we can calculate the composition of  $\tilde{x}$  and  $\tilde{y}$ , and then compose with  $\tilde{z}$ . The later, in combination with Proposition 3, implies  $u \sim (xy)z$ .  $\square$

There is also a correspondence for the right-distributive law.

**Proposition 5.** *If  $w \sim (x + y)z$ , then  $\tilde{w}(\theta) = \tilde{x}(\tilde{z}(\theta)) + \tilde{y}(\tilde{z}(\theta))$ .*

*Proof.* This follows directly from the previous results and the rules of function composition.  $\square$

We can use the Taylor expansions of CGFs to derive some general rules for the calculation of cumulants of sums and products of NCRVs, for instance. If

$$\tilde{x}(\theta) = a_1\theta + \frac{a_2}{2}\theta^2 + O(\theta^3), \quad (19)$$

$$\tilde{y}(\theta) = b_1\theta + \frac{b_2}{2}\theta^2 + O(\theta^3), \quad (20)$$

then

$$\tilde{x}(\theta) + \tilde{y}(\theta) = (a_1 + b_1)\theta + \left(\frac{a_2 + b_2}{2}\right)\theta^2 + O(\theta^3), \quad (21)$$

and

$$\begin{aligned} \tilde{x}(\tilde{y}(\theta)) &= a_1 \left[ b_1\theta + \frac{b_2\theta^2}{2} + O(\theta^3) \right] + \frac{a_2}{2} \left[ b_1\theta + \frac{b_2\theta^2}{2} + O(\theta^3) \right]^2 + O(\theta^3) \\ &= a_1b_1\theta + \left( \frac{a_1b_2 + a_2b_1^2}{2} \right) \theta^2 + O(\theta^3). \end{aligned} \quad (22)$$

Thus,

$$\langle x + y \rangle = \langle x \rangle + \langle y \rangle, \quad \text{and} \quad \langle\langle x + y \rangle\rangle = \langle\langle x \rangle\rangle + \langle\langle y \rangle\rangle, \quad (23)$$

while

$$\langle xy \rangle = \langle x \rangle \langle y \rangle, \quad \text{and} \quad \langle \langle xy \rangle \rangle = \langle x \rangle \langle \langle y \rangle \rangle + \langle \langle x \rangle \rangle \langle y \rangle^2. \quad (24)$$

We see that expectations behave commutatively, since  $\langle xy \rangle = \langle yx \rangle$ , but variances do not, since in general  $\langle \langle xy \rangle \rangle \neq \langle \langle yx \rangle \rangle$ . The result that the expectation of the products of NCRVs is the product of their expectations is a special case of Wald's equation.

We can also use the correspondence between NCRVs and their CGFs to directly derive CGFs for new random variables. Observe that the CGF for a constant integer  $n$  is  $\tilde{n}(\theta) = n\theta$ , and the CGF for a Bernoulli random variable  $b$  that returns 1 with probability  $p$  and 0 with probability  $1 - p$  is  $\tilde{b}(\theta) = \ln(1 - p + pe^\theta)$ . Now, suppose we want to construct a new variable  $x$  that is the sum of  $n$  independent Bernoulli variables. Then  $x \sim nb$ , so  $\tilde{x}(\theta) = \tilde{n}(\tilde{b}(\theta)) = n \ln(1 - p + pe^\theta)$ . In fact, this is the CGF of a binomial random variable with parameters  $p$  and  $n$ . Similarly, the negative binomial distribution is the sum of  $n$  geometrically distributed random variables. Table 1 lists cumulant generating functions for some common distributions.

In epidemiology, random variable multiplication is a natural way to model the transmission of an infectious disease. Suppose  $x$  is a random variable describing the number times a sick individual transmits a disease to susceptible individuals. If each of these newly infected individual also transmits the disease a random number  $x$  times, then the second generation of transmission will have  $xx \sim x^2$  sick people. Similarly,  $n$ 'th generation of transmission will have  $x^n$  sick individuals. If  $x$  has a CGF  $\tilde{x}(\theta)$ , then the CFG of  $x^n$  will be an  $n$ -fold composition of  $\tilde{x}$ .

Suppose, instead of just needing to know the number of individuals sick in the  $n$ 'th generation, we actually want to know the total number of individual who become sick over the course of the epidemic. We can use the random-variable algebra to derive an equation for this CGF as well. Without loss of generality, we can assume the epidemic starts with 1 sick individual. Let  $T(t)$  be a random variable for the total number of sick individuals caused by one infection after  $t$  generations of transmission. Then  $T(1) = 1$  and the total number of sick individuals after  $t + 1$  generations of transmission is equal to the first individual plus all the individuals infected in  $t$  generations for each infection caused by the first individual,  $T(t + 1) \sim 1 + xT(t)$ . Taking the limit as  $t \rightarrow \infty$ ,

$$T(\infty) \sim 1 + x(1 + x(1 + x(\dots))) \sim 1 + xT(\infty). \quad (25)$$

Note that, since multiplication is not right-distributive,

$$T(\infty) \approx \sum_{t=0}^{\infty} x^t. \quad (26)$$

Table 1: Common cumulant generating functions

Name	Probability Distribution	Cumulant Generator
Generic Integer Distribution	$f_x : x \in \mathbb{Z}$	$\tilde{f}(\theta) = \ln \sum_{x \in \mathbb{Z}} f_x e^{\theta x}$
Bernoulli	$p^x(1-p)^{1-x}, x \in \{0, 1\}$	$\ln(1-p+pe^\theta)$
Binomial	$\frac{n!}{x!(n-x)!} p^x(1-p)^{n-x}, x \in \{0 \dots n\}$	$n \ln(1-p+pe^\theta)$
Poisson	$\frac{\lambda^x}{x!} e^{-\lambda}$	$(e^\theta - 1) \lambda$
Geometric	$(1-p)p^x$	$\ln(1-p) - \ln(1-pe^\theta)$
Negative Binomial	$\frac{\Gamma(a+x)}{\Gamma(a)\Gamma(x+1)} (1-p)^a p^x$	$a [\ln(1-p) - \ln(1-pe^\theta)]$
Multitype Bernoulli	$\mathbf{p}^{\mathbf{x}}$	$\ln(\sum_i p_i e^{\theta_i})$
Multinomial	$\frac{\mathbf{n}!}{\mathbf{x}!} \mathbf{p}^{\mathbf{x}}$	$n \ln(\sum_i p_i e^{\theta_i})$
Point mass	$\delta(x-n)$	$n\theta$
Uniform	$\frac{I_{[a,b]}(x)}{b-a}$	$\ln \frac{e^{b\theta} - e^{a\theta}}{(b-a)\theta}$
Exponential	$\lambda e^{-\lambda x}$	$\ln \sum_{j=0}^{\infty} \left(\frac{\theta}{\lambda}\right)^j \approx -\ln 1 - \frac{\theta}{\lambda}$
Gamma	$\lambda (\lambda x)^{c-1} \frac{e^{-\lambda x}}{\Gamma(c)}$	$-c \ln 1 - \frac{\theta}{\lambda}?$
Normal	$\frac{e^{-(x-m)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$	$\frac{1}{2}\sigma^2\theta^2 + m\theta$
Cauchy	$\frac{1}{\pi} \frac{\alpha}{\alpha^2 + (x-m)^2}$	$m\theta - \alpha \theta/i $ (formal)
Laplace	$\frac{\lambda}{2} e^{-\lambda x }$	$\ln \frac{1}{2} I_{[-\lambda, \lambda]} \left[ \left(\frac{\theta}{\lambda} - 1\right)^{-1} + \left(\frac{\theta}{\lambda} + 1\right)^{-1} \right]$
Bessel	$\frac{\lambda}{D\sqrt{\pi}\Gamma(n)} \sqrt{\left(\frac{\lambda x^2}{4D}\right)^{n-1/2}} K_{n-1/2}\left(\sqrt{\frac{\lambda x^2}{D}}\right)$	$-n \ln 1 - \frac{D}{\lambda} \theta^2$

By the correspondence principle, if  $\tilde{T}(\infty; \theta)$  is the CGF of  $T(\infty)$  then

$$\tilde{T}(\infty; \theta) = \theta + \tilde{x} \left( \tilde{T}(\infty; \theta) \right). \quad (27)$$

$\tilde{T}(\infty; \theta)$  can then be calculated by solving this equation algebraically or using series methods. For instance, if a sick person infects two new individuals with probability  $p$  and no individuals with probability  $1 - p$ , then

$$\tilde{x}(\theta) = \ln \left( 1 - p + pe^{2\theta} \right) \quad (28)$$

and

$$\tilde{T}(\infty; \theta) = \theta + \ln \left( 1 - p + pe^{2\tilde{T}(\infty; \theta)} \right). \quad (29)$$

Solving for  $\tilde{T}(\infty; \theta)$ ,

$$\tilde{T}(\infty; \theta) = \ln \left[ 1 - \sqrt{1 - 4p(1 - p)e^\theta} \right] - \ln 2p. \quad (30)$$

We can calculate the probability that the epidemic goes extinct, which corresponds to the probability that the epidemic size is finite. For any CGF  $\tilde{y}$ ,  $\tilde{y}(\theta = 0) = \ln \langle 1 \rangle$ , where  $\langle 1 \rangle$  is the sum of the probabilities of all *finite* states. Applied to Eq. (30),

$$e^{\tilde{T}(\infty; 0)} = \frac{\left[ 1 - \sqrt{1 - 4p(1 - p)} \right]}{2p} \quad (31)$$

is the probability that the epidemic eventually goes extinct and has a finite number of cases. This is also the answer to the classic problem of the extinction of surnames studied by Galton and Watson.

The reader may wonder at this point if these methods can be extended beyond the positive integers to other number systems. The answer is a qualified yes, but not with the full generality we might naively hope for. In particular, there is no universal way to extend left-multiplication by a negative, rational, or real number, since we can not have negative realizations of a variable or a fraction of a realization. With care, however, useful results can be reconstructed. For instance, we can usually make sense of right-multiplication by arbitrary-valued random variables over which we can define an addition operation. As one example, NCRVs where we define right-multiplication by a real number can be used to motivate the central limit theorem. Suppose  $x$  is a real-valued NCRV,  $n$  is an integer-valued deterministic variable. As

classically formulated [Feller, 1968], the central limit theorem states that if  $y$  is a sequence of random variables with

$$y \sim \frac{x_1 + x_2 + \dots + x_n}{n}, \quad (32)$$

then as  $n \rightarrow \infty$ , the distribution of  $y$  converges to a Gauss distribution with the same mean as  $x$  and a variance  $\langle\langle y \rangle\rangle = \langle\langle x \rangle\rangle/n$ . To motivate this using NCRV algebra, we first note that if  $\lambda$  is a positive real number describing a dilation of  $x$ , then the CGF of  $x$  times  $\lambda$  is

$$\ln \langle e^{\theta x \lambda} \rangle = \ln \langle e^{(\theta \lambda) x} \rangle = \tilde{x}(\lambda \theta). \quad (33)$$

If we take  $n$  to be the deterministic integer number of samples of  $x$  and  $n^{-1}$  to be the deterministic amount we shrink each sample by, then

$$y \sim n x n^{-1}. \quad (34)$$

If  $x$  has CGF

$$\tilde{x}(\theta) = \langle x \rangle \theta + \frac{\langle\langle x \rangle\rangle}{2} \theta^2 + O(\theta^3) \quad (35)$$

and we explicitly calculate the CGF of  $y$ ,

$$\tilde{y}_n(\theta) = n \left[ \langle x \rangle n^{-1} \theta + \frac{\langle\langle x \rangle\rangle}{2} (n^{-1} \theta)^2 + O(n^{-3} \theta^3) \right] \quad (36)$$

$$= \langle x \rangle \theta + \frac{\langle\langle x \rangle\rangle}{2n} \theta^2 + O(n^{-2} \theta^3) \quad (37)$$

By inspection, we see that the CGF of  $y$  converges to that of a Gauss distribution with expectation  $\langle x \rangle$  and variance  $\langle\langle x \rangle\rangle/n$  for large  $n$ . Of course, this is not a fully general result because  $x$  is restricted to integer values, but it is a useful intermediate step for students.

As a second example of working with non-integer valued random variables, suppose a particle diffuses along a line from an initial position at the origin, but has a constant hazard of stopping at any given instant. What is the distribution of settling positions? This problem was originally solved in an ecology context by Yamamura [2002], where the results were obtained using standard methods. Here, we show how a more general result is obtained using NCRVs .

A particle is released at the origin at time 0. This particle advects and diffuses along a line for some random time before settling in its final location. If the hazard

of settling in any given instant is constant, then for a small time interval  $\Delta t$  the particle settles with a small probability  $\lambda\Delta t$ . Let the number of time steps before a particle settles be represented by an integer-valued NCRV  $N_{\Delta t}$  that is geometrically distributed with CGF

$$\tilde{N}_{\Delta t}(\theta) = \ln \left( \frac{\lambda\Delta t}{1 - (1 - \lambda\Delta t)e^\theta} \right). \quad (38)$$

Now, over the same small time interval  $\Delta t$ , the displacement  $s_{\Delta t}$  of an advecting and diffusing particle is normally distributed like

$$\frac{1}{\sqrt{4\pi D\Delta t}} e^{-\frac{(x-a\Delta t)^2}{4D\Delta t}} \quad (39)$$

where  $a$  is the advection rate and  $D$  is the diffusion rate. The CGF of the displacement is

$$\tilde{s}_{\Delta t}(\theta) = a\Delta t\theta + D\Delta t\theta^2. \quad (40)$$

Having descriptions of both the movement and settling processes, the final position of the particle is approximately the sum of the displacements over each interval before the particle settles. Since the number of intervals is given by the NCRV  $N_{\Delta t}$ , and all displacements are independent and identically distributed, then the final position is approximately given by the NCRV  $y_{\Delta t} \sim N_{\Delta t}s_{\Delta t}$ , and the corresponding CGF

$$\tilde{y}_{\Delta t}(\theta) = \ln \left( \frac{\lambda\Delta t}{1 - (1 - \lambda)e^{a\Delta t\theta + D\Delta t\theta^2}} \right). \quad (41)$$

Taking the limit as the time interval  $\Delta t$  becomes infinitesimal,

$$\lim_{\Delta t \rightarrow 0} \tilde{y}_{\Delta t}(\theta) = -\ln \left( 1 - \frac{a\theta + D\theta^2}{\lambda} \right). \quad (42)$$

In the special case of  $a = 0$ , standard tables show that this is the CGF of a Bessel distribution. Thus, the settling distribution for a diffusing particle with no advection and a constant risk of settling is a Bessel distribution. Our more general result, however, can be used in other calculations like that of determining spreading speeds of invasive species. Similar results can be derived in any case where an appropriate  $\Delta t$ -scaling of the settlement time can be constructed.

An interesting aspect of this derivation is that it involves two continuous NCRVs, one for time and one for space. In order to capture multiplication, we had to discretize time and look at the limit of small time steps, but were able to leave space as a continuous variable.

## 5 Continuous-Time Autonomous Processes

Small time steps can also be used to derive differential equations describing the continuous change in a process's state. Suppose that the NCRV  $x_{\Delta t}$  represents one individual's contribution to the change of a branching process's state over a small window of time  $\Delta t$ . Then over a time window  $n\Delta t$ , the cumulative change in the population is described by the NCRV equation

$$\Xi(n\Delta t) \sim \Xi(0)x_{\Delta t}^n, \quad (43)$$

where  $\Xi(0)$  represents the population's initial state at time  $t = 0$ . Progressing one more increment  $\Delta t$  forward in time,

$$\Xi(n\Delta t + \Delta t) \sim \Xi(0)x_{\Delta t}^n x_{\Delta t} \sim \Xi(n\Delta t)x_{\Delta t}. \quad (44)$$

These NCRV equations can be used to obtain differential equations for the evolution of the CGF. As a simple example, consider a density-independent birth-death process. The probability of an event happening in a small window of time  $\Delta t$  is  $\lambda\Delta t$ . That event is a birth with probability  $p$  and a death with probability  $1 - p$ . It follows that the CGF representing the change to an individual over this small window  $\Delta t$  is

$$\tilde{x}_{\Delta t}(\theta) = \ln [(1 - \lambda\Delta t)e^\theta + p\lambda\Delta te^{2\theta} + (1 - p)\lambda\Delta t] \approx \theta + [pe^{2\theta} - e^\theta + (1 - p)] \lambda\Delta t \quad (45)$$

The coefficient of  $\Delta t$  in the series approximation is called the infinitesimal generator of the process. Autonomous branching processes can be identified completely by their generators. If  $\Xi(t)$  is the NCRV representing the state of the population at time  $t$ , then the CGF for the random variable  $\Xi(t + \Delta t)$  representing state of the population at time  $t + \Delta t$  is

$$\tilde{\Xi}(t + \Delta t, \theta) = \tilde{\Xi}(t, \theta + [pe^{2\theta} - e^\theta + (1 - p)] \lambda\Delta t). \quad (46)$$

In the limit of  $\Delta t \rightarrow 0$ ,

$$\frac{\partial \tilde{\Xi}(t, \theta)}{\partial t} = \frac{\partial \tilde{\Xi}(t, \theta)}{\partial \theta} \lambda [pe^{2\theta} - e^\theta + (1 - p)]. \quad (47)$$

So the change in the CGF is described by a first-order linear partial differential equation. This equation is referred to as the forward equation because it describes the evolution of an arbitrary distribution of states forward in time. From this partial differential equation, we can obtain simpler equations for the mean, variance, and extinction probabilities.

In addition to the forward equation, the infinitesimal generator also defines a backward equation which describes the CGFs for a process starting with exactly one individual at time  $t = 0$ . Since (45) is the CGF for the first increment of time for the population's change,

$$\tilde{x}(t + \Delta t) = \tilde{x}(t) + [pe^{2\tilde{x}(t)} - e^{\tilde{x}(t)} + (1 - p)] \lambda \Delta t. \quad (48)$$

In the infinitesimal limit, we find the backward equation defining the CGF is

$$\frac{d\tilde{x}}{dt} = \lambda [pe^{2\tilde{x}} - e^{\tilde{x}} + (1 - p)] \quad (49)$$

with initial condition  $\tilde{x}(t = 0, \theta) := \theta$ .

By differentiating the equations for the CGFs, we can get equations for the dynamics of the mean and variance over time. First derivatives of both Eq. (47) and Eq. (49) with respect to  $\theta$  and evaluated at  $\theta = 0$  yield

$$\frac{1}{\lambda} \frac{d \langle x \rangle}{dt} = (2p - 1) \langle x \rangle. \quad (50)$$

But the variance equations are different in the forward and backward cases. Forward in time, the variance evolves according to the linear equation

$$\frac{1}{\lambda} \frac{d \langle \langle x \rangle \rangle}{dt} = (2p - 1) \langle \langle x \rangle \rangle + (4p - 1) \langle x \rangle \quad (51)$$

Backward in time, the variance evolves according to the nonlinear equation

$$\frac{1}{\lambda} \frac{d \langle \langle x \rangle \rangle}{dt} = (2p - 1) \langle \langle x \rangle \rangle + (4p - 1) \langle x \rangle^2 \quad (52)$$

For further discussion of related models, see Matis and Kiffe [2000].

## 6 Conclusion

The basic method of studying branching processes using generating functions was introduced more than a century ago, and is used frequently throughout mathematics. Here, I have described an alternative perspective on branching processes using non-commuting random variable algebra and cumulant generating functions. This approach simplifies some of the mathematical machinery usually needed in manipulation of branching processes. The connections between modelling concepts,

mathematical formulations, and simulation algorithms are center-stage, and can be grasped quickly. This approach leverages people’s common knowledge of algebra and statistics. Modelling with branching processes is reduced to constructing some probability distributions, writing some standard algebra equations, and applying the rules of function composition. Results are easily derived in terms of both means and variances. In addition, the NCRV approach is another example of how algebra laws like the commutative law and the distributive law are special properties of some systems rather than universal laws.

Our discussion of cumulant generating functions has been concerned mostly with single-type variables. Useful results for multi-type NCRVs can be found by extending the definitions of generating functions to include multiple variables:

$$\tilde{F}(\theta_1, \theta_2, \dots) = \ln \langle e^{F_1\theta_1 + F_2\theta_2 + \dots} \rangle.$$

In this way, covariance and conditional extinction relationships can be derived. Similarly, generating functions for continuously-typed processes can be constructed using functional integration methods. This leads to a particularly nice approach to the Fokker-Plank equations for diffusion.

The practical use of the NCRV approach remains unclear. Perhaps they are no more than a convenient trick. All of the examples described here, with the possible exception of parts of Eq. (42), are standard. They are closely related to the geometric rewrite rules used in plant morphology [Prusinkiewicz and Lindenmayer, 1990, Meinhardt, 2003]. The extensions to multitype and continuous-type processes allow for the study of the Fokker-Plank equation and related processes. Certainly, mathematicians have powerful tools for the study of branching processes that do not rely on the non-commuting random variable framework. But sometimes alternative perspectives are fruitful, making complicated arguments simple and suggesting new avenues of research.

## Acknowledgements

The author would like to thank D. Brian Walton and two anonymous reviewers for very helpful comments on earlier versions of this manuscript.

## References

K. B. Athreya and P. E. Ney. *Branching Processes*. Springer-Verlag, New York, NY, 1972.

- Hal Caswell. *Matrix Population Models*. Sinauer, Sunderland, MA, 2nd edition, 2001.
- Karin S. Dorman, Janet S. Sinsheimer, and Kenneth Lange. In the garden of branching processes. *SIAM Review*, 46(2):202–229, 2004.
- William Feller. *An introduction to probability theory and its applications*, volume 1. Wiley, New York, NY, 2nd edition, 1968.
- Francis Galton. *Natural Inheritance*. Macmillan, 2nd edition, 1889.
- Patsy Haccou, Peter Jagers, and Vladimir Alekseevich Vatutin. *Branching processes: variation, growth, and extinction of populations*. Cambridge University Press, 2005.
- Theodore E. Harris. *The Theory of Branching Processes*. Springer–Verlag, 1963.
- Marek Kimmel and David E. Axelrod. *Branching processes in biology*. Springer, New York, 2002.
- C. Loi and P.H. Cournède. Generating functions of stochastic L-systems and application to models of plant development. *DMTCS Proceedings*, (1), 2008.
- C. A. Macken and A. S. Perelson. *Branching Processes Applied to Cell Surface Aggregation Phenomena*, volume 58 of *Lecture Notes in Biomathematics*. Springer-Verlag, New York, NY, 1985.
- James H. Matis and Thomas Kiffe. *Stochastic Population Models: A Compartmental Perspective*. Springer-Verlag Telos, New York, NY, 2000.
- H. Meinhardt. *The Algorithmic Beauty of Sea Shells*. Springer, 2003.
- R. Meza, J. Jeon, S.H. Moolgavkar, and E.G. Luebeck. Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proceedings of the National Academy of Sciences*, 105(42):16284, 2008.
- Charles J. Mode. *Multitype Branching Processes: Theory and Applications*. American Elsevier, New York, NY, 1971.
- MEJ Newman, SH Strogatz, and DJ Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118, 2001.
- I. Newton. *Philosophiae Naturalis Principia Mathematica*. London, 1686.

- P. Prusinkiewicz and A. Lindenmayer. *The algorithmic beauty of plants*. Springer-Verlag, New York, NY, 1990.
- P. Prusinkiewicz, Y. Erasmus, B. Lane, L.D. Harder, and E. Coen. Evolution and development of inflorescence architectures. *Science*, 316(5830):1452, 2007.
- T. C. Reluga, R. Meza, D. B. Walton, and A. P. Galvani. Reservoir interactions and disease emergence. *Theoretical Population Biology*, 72(3):400–408, 2007.
- Stephen M. Stigler. *The History of Statistics: The Measurement of Uncertainty before 1900*. Harvard University Press, Cambridge, MA, 1986.
- S. M. Ulam. *Analogies Between Analogies: The mathematical reports of S. M. Ulam and his Los Alamos collaborators*. Los Alamos Series in Basic and Applied Sciences. University of California Press, 1990.
- Herbert S. Wilf. *Generatingfunctionology*. Academic Press, New York, NY, 1999.
- David Williams. *Probability with Martingales*. Cambridge University Press, 1991.
- Kohji Yamamura. Dispersal distance of heterogeneous populations. *Population Ecology*, 44:93–101, 2002.