

# WHEN DID JOE'S GREAT . . . GRANDFATHER LIVE? OR: ON THE TIME SCALE OF EVOLUTION.

Peter Jagers, Olle Nerman, and Ziad Taib  
Chalmers University of Technology and Gothenburg University

## Abstract

In a general supercritical branching process the distribution is found of the time back to the birth of the  $n$ :th father of a random individual. For single-type processes a probability of mutation (at a site of a specified gene) is introduced. With the infinite alleles interpretation that every mutation occurring is unique the class of individuals with the same allele can be viewed as a generalized "macroindividual". Thereby the question of when the first of a series of  $n$  mutations occurred is reduced to finding the birth time of the  $n$ :th father in the process of macroindividuals. This is done.

**1. Introduction** The three great problems of population dynamics are the extinction, the growth, and the composition of populations. The first makes sense already in the simplest of branching processes, the Galton-Watson model, which was indeed born out of it. The second requires a physical time, not only generation counting, and for the third problem one should have access also to the birth order of siblings and times between their births, and preferably a possibility to discern between individuals of different inherited types.

This leads to general, multi-type branching processes (Jagers and Nerman 1984 (the one-type case), Nerman 1984, Jagers 1989 (abstract type spaces)). When an individual is born it inherits a *type*,  $s$ , from a *type space*,  $S$  with a  $\sigma$ -algebra  $\mathcal{S}$ . The type, which you may think of as a genotype, determines a probability kernel,  $P(s, \cdot)$ , over a *life space*,  $(\Omega, \mathcal{A})$ , of all possible life careers. The life space may be very rich in order to contain all relevant aspects of individual life, and is best thought of as an abstract measurable space. What has to be defined on it - otherwise there would be no population to talk of - is a *reproduction process* for individuals. It will be denoted by  $\xi$  and viewed as a point process on  $S \times \mathbf{R}_+$ , the first coordinate of any point yielding the type of that child, the second the father's age at the birth<sup>1</sup>. The processes is supposed finite on bounded subsets of  $\mathbf{R}_+$ , so that the children can be numbered in birth order (arbitrarily for twins). If  $\sigma_k, \tau_k$  are thus the type of the  $k$ :th child and father's age at his birth, then

$$\xi(A \times B) = \# \{k; \sigma_k \in A, \tau_k \in B\}, A \in \mathcal{A}, B \in \mathcal{B},$$

$\mathcal{B}$  denoting the Borel algebra on  $\mathbb{R}_+$ . A biologically basic, though not mathematically fundamental, entity defined on the life space is the *life span*,

$$\lambda : \Omega \rightarrow \bar{\mathbb{R}}_+,$$

the extended non-negative half-line.

The individuals to lead these lives are denoted in the classical Ulam-Harris fashion: the ancestor is 0, the first generation is  $\{1,2,\dots\} = N$ , the  $n$ :th is  $N^n, N^0 = \{0\}$ , and the set of all possible individuals is written

$$I = \bigcup_{n \geq 0} N^n.$$

If individuals and sets of individuals are partially ordered by descent ( $x$  stems from  $y$  if  $y = 0$  or  $y$  is the first part of  $x$ ,  $A$  stems from  $B \subset I$  if for all  $x \in A$  there is some  $y \in B$  such that  $x$  stems from  $y$ ), then the *life kernel*  $P(s, \cdot)$ ,  $s \in S$  determines a Markov field over sets of individuals (Jagers 1989) and the flow of time via the following dynamics:

- The ancestor 0 is born at time 0 and of an arbitrary type  $\sigma_0 = s \in S$ .
- He gets  $\xi_0(S \times \mathbb{R}_+)$  children, viz.  $1,2,\dots, \xi_0(S \times \mathbb{R}_+)$ , at times  $\tau_1, \tau_2, \dots, \tau_{\xi_0(S \times \mathbb{R}_+)}$  and they are of types  $\sigma_1, \sigma_2, \dots, \sigma_{\xi_0(S \times \mathbb{R}_+)}$ . These are measurable functions of the ancestor's life, chosen from the life space  $\Omega$  according to  $P(\sigma_0, \cdot)$ .
- The individual  $x = (x_1, x_2, \dots, x_n)$  ( $= x_1 x_2 \dots x_n$  for short),  $x_i \in N$ , is the  $x_n$ :th child of  $x_1 x_2 \dots x_{n-1}$ . He is of type  $\sigma_x$  and is born at time  $\tau_x$  which is the sum of 0's age  $\tau_{x_1}$  at begetting  $x_1$ , the latter's age at giving birth to his  $x_2$ :th child *etc.*. The interpretation of  $\tau_x = \infty$  is, of course, that  $x$  is never born.

The probability measure thus arising on the *population space*  $(S \times \Omega^I, S \times \mathcal{A}^I)$  will be denoted by  $P_s$  for  $\sigma_0 = s \in S$  with expectation operator  $E_s$ . If the starting type was chosen according to some distribution, say  $\pi$ , over  $S$ . We write  $P_\pi$  and  $E_\pi$ .

A couple of rather technical assumptions (cf. Jagers 1989) are needed. They can be summarized by saying that the process should be

- Malthusian, *i.e.* there is a Malthusian growth parameter  $\alpha$ ,

- supercritical, *i.e.*  $0 < \alpha < \infty$ ,
- irreducible, *i.e.* individuals of almost all (with respect to a *stable type distribution*  $\pi$ , see beneath) types can have descendants of almost all types,
- strongly  $\alpha$ -recurrent, meaning that the mean age at childbearing in the population is (positive and) finite, and
- spread-out in the sense that the *reproduction kernel*

$$\mu(s, A \times B) = E_s[\xi(A \times B)] = \int_{\Omega} \xi(A \times B)(\omega) P(s, d\omega)$$

is so with respect to the product of the stable type and Lebesgue measures.

The strict definition of the stable type measure  $\pi$  is that it is the maximal measure satisfying

$$\int_0^{\infty} e^{-\alpha t} \mu(\pi, A \times dt) = \pi(A) \cdot \mu(\pi, A \times B) = \int_S \mu(s, A \times B) \pi(ds),$$

It is assumed normed to a probability measure.

Now assume that there is given a so called *random characteristic* *i.e.* a weight, that for each individual might be influenced by his type, age, life and progeny, but not of his ancestors' lives. (Jagers 1975 gives many examples, one will follow here.) Make the dependence upon age explicit, writing the characteristic  $\chi(a)$ ,  $a$  for age. At time  $t$  sample an individual at random from among all those born (*the total population*) or from among all those alive at that time (or possibly from some other subset of the total population). Call that individual *Joe* and check his  $\chi$ -value. The population development given, its expectation will be the average  $\chi$ -value in the population now at time  $t$ . Then, let  $t \rightarrow \infty$ . Will this conditional expected  $\chi$ -value stabilize? Under certain technical conditions (including the famous *xlogx*) the answer is "yes". The limit, which we shall call simply Joe's expected  $\chi$ -value in the stable case, is given by the following theorem, stated quite loosely here. (For a precise formulation and proof *cf.* Jagers 1989.) In it, and beneath, *AGE* is an exponentially distributed random variable with parameter  $\alpha$  independent of everything else, the index  $\pi$  indicates that the expectation is taken with respect to a branching process whose ancestor had a type, random according to the stable type distribution,  $\chi$  and  $\lambda$  refer to this ancestor, and the hat to Laplace transform,  $\hat{\chi}(\alpha) = \alpha \int_0^{\infty} e^{-\alpha t} \chi(t) dt$ .

**Theorem 1.** *In a stable population Joe's expected  $\chi$ -value is  $E_{\pi}[\hat{\chi}(\alpha)]$  if he was sampled from the total population,  $E_{\pi}[\hat{\chi}(\alpha) | AGE > \lambda]$  if sampled from among those alive.*

Famous special cases of this theorem are the stable age distribution of demography (the characteristic being the indicator of age intervals) and the mitotic index of cell kinetics (where the characteristic is the indicator of the cell cycle time in mitosis). The theorem is itself a special case of a description of the stable population generally, *i.e.* also for characteristics influenced by Joe's progenitors, cousins *etc.*, *cf.* Jagers and Nerman 1984 for the one-type case and Nerman 1984 for the general theory. Since Nerman's paper is not so easily available we shall, however use the theorem in the next section. For the one-type case *cf.*, however, Nerman and Jagers 1984. But certainly the general stable population provides the proper context for the problem to be treated.

**2. When Was Joe's  $n$ :th Father Born?** What is, thus, the probability that Joe's  $n$ :th father, *i.e.* Joe's father's father's ... father ( $n$  times) was born no more than  $u$  time units ago? To make use of the theorem we must find a characteristic that counts an individual if his  $n$ :th father satisfies this. Since it may not be influenced by ancestors of the individual it measures, we have to refer it to the  $n$ :th father himself, giving him the weight  $k$  if he has  $n$ :th generation descendants within time  $u$ :

$$\chi^u(a) = \sum_{n \in N^n} 1_{\{\tau_x \leq a \leq u\}}.$$

(In this first step we do not bother whether the  $n$ :th generation descendant is alive or not.) Clearly

$$\hat{\chi}^u(a) = \sum_{n \in N^n} \alpha \int_{\tau_x}^u e^{-\alpha a} da 1_{\{\tau_x \leq u\}} = \sum_{n \in N^n} (e^{-\alpha \tau_x} - e^{-\alpha u}) 1_{\{\tau_x \leq u\}}.$$

For functions of a non-negative real variable let  $*$  denote convolution, and for functions on  $S \times \mathbf{R}_+$  let it stand for transition on  $S$  and convolution on  $\mathbf{R}_+$ . Thus,

$$\mu^{*2}(A \times B) = \mu * \mu(A \times B) = \int_{S \times \mathbf{R}_+} \mu(r, A \times B - t) \mu(s, dr \times dt)$$

Also write

$$\mu_{\alpha}(s, dr \times dt) = e^{-\alpha t} \mu(s, dr \times dt).$$

Since by induction

$$\sum_{x \in N^n} P_\pi(\tau_x \leq u) = \mu^{*n}(\pi, S \times [0, u]),$$

it follows that

$$\mathbb{E}_\pi[\hat{\chi}^\mu(\alpha)] = \int_0^\mu (e^{-\alpha t} - e^{-\alpha u}) \mu^{*n}(\pi, S \times dt) = E_\alpha * \mu_\alpha^{*n}(\pi, S \times [0, \mu]),$$

$E_\alpha$  denoting the exponential distribution with parameter  $\alpha$ .

The case of proper interest, Joe alive, is somewhat more difficult in this setting. (It is easy if you have the general stable population theory.) We cannot apply the conditional version of the theorem, since the characteristic has been carried back to referring to Joe's  $n$ :th father, rendering him the Joe of the theorem. Instead we have to consider a characteristic

$$\chi_l^\mu(a) = \sum_{n \in N^n} 1_{\{\tau_x \leq a \leq u, \lambda_x > a - r_x\}},$$

the index  $x$  indicating that not only the birth-time refers to the individual  $x$  but also the life-span. Calculating as above leads to an expression involving, besides the reproduction function, also the *life span distribution*,  $L$  of an individual whose type is determined according to the stable type distribution,  $\pi$ ,

$$\begin{aligned} \mathbb{E}_\pi[\hat{\chi}_l^\mu(a)] &= \mathbb{E}_\pi \left[ \sum_{n \in N^n} \alpha \int_{\tau_x}^{\mu \wedge (\tau_x + \lambda_x)} e^{-\alpha a} da 1_{\{\tau_x \leq u\}} \right] = \dots = \\ &= \int_0^\mu \left( \int_0^{(\mu-t)} e^{-\alpha v} (1 - L(v)) dv \right) \mu^{*n}(\pi, S \times dt). \end{aligned}$$

This is, however, not the correct expression: we have to norm by the probability of being alive,

$$\int_0^\infty e^{-\alpha t} (1 - L(t)) dt.$$

Then it follows that the probability of Joe's  $n$ :th father being born no more than time  $\mu$  ago has the form

$$L_\alpha * \mu_\alpha^{*n}(\pi, S \times [0, \mu]),$$

$L_\alpha$  being the stable age distribution

$$\frac{\int_0^\mu e^{-\alpha t} (1 - L(t)) dt}{\int_0^\infty e^{-\alpha t} (1 - L(t)) dt}.$$

If life length is infinite, this reduces to the exponential distribution, of course.

We specialize now to the one-type case where there is a proper reference for the stable population derivation of this result, Nerman and Jagers 1984. Then  $\mu(\pi, S \times [0, u])$  is merely the *reproduction function*  $\mu(u)$  and  $\mu_\alpha(\pi, S \times [0, u])$  a probability measure  $\mu_\alpha$  on  $\mathbf{R}_+$ , the stable distribution of age at childbearing. Its expectation

$$\beta = \int_0^\infty t \mu_\alpha(dt) = \int_0^\infty t e^{-\alpha t} \mu(dt)$$

is the *mean age at childbearing*. Its variance will be denoted by  $v^2$ .

The central limit theorem thus yields that the time back to Joe's  $n$ :th father's birth is, approximately,

$$X + n\beta + v\sqrt{n}N(0, 1),$$

$X$  being exponential ( $\alpha$ ) if Joe was sampled from the total population and following the stable age distribution, if he was chosen from those alive, and independent of the standard normal variable  $N(0,1)$  in both cases. If there is further information about Joe this will reflect itself only on  $X$ . *E.g.* if Joe is known to be 65, the time back will be

$$65 + n\beta + v\sqrt{n}N(0, 1).$$

**Example.** In the last mentioned case we find out when Joe's greatgreatgreatgreatgreatgreatgrandfather was born. Assume that  $\beta = 30$  and  $v = 7$ . An approximately 95% prediction (or, rather, postdiction) interval is then

$$1989 - 65 - 300 \pm 2 \times 7 \times \sqrt{10} \approx 1624 \pm 46 = [1578, 1670].$$

**3. The Time Scale of Evolution.** Through the advent of molecular biology it has become possible to study the process of evolution at the level of nucleic and amino acid sequences. There is a host of literature, *cf.* Kimura 1979, Wilson *et al.* 1987, or in a branching process context Taib 1987. Much of the statistical-ly oriented work, *cf.* Barry and Hartigan 1987, is concerned with the reconstruction of the evolutionary tree. We wish to contribute to the understanding of the time scale in given trees.

Thus consider some gene or protein, like the often studied  $\alpha$ - or  $\beta$ -globins, and assume that there is, at each birth a probability  $p$  of mutation. In accordance with the infinite alleles interpretation each mutation is supposed to be unique. The underlying population is viewed as a supercritical, one-type branching process with the reproduction function  $\mu$ , as in the preceding section. The expected number of first mutant progeny of an individual, *i.e.* of mutant individuals stem-

ming from the individual and being the first mutants in their line of descent, is then

$$M(t) = p \sum_{n=1}^{\infty} q^{n-1} \mu^{*n}(t), \quad q = 1-p,$$

within time  $t$ . Following Taib 1987, we clump an individual and all his non-mutant progeny into a *macroindividual*, born at the birth of the first individual of that allelic type, dead at the death of the last non-mutant descendant.

This creates a new branching process of macroindividuals, having the reproduction function  $M$ . The question how long a time ago the first in a sequel of  $n$  mutations occurred in the original population reduces to the question about Joe's  $n$ :th father in the population of macroindividuals, *cf.* also the more general analysis in Taib 1987, Chapter 7. To apply the answer of the latter to the former question it remains only to determine the Malthusian rate and the expectation and variance of the new process. But this is easy:

$$\hat{M}(\alpha) = p \sum_{n=1}^{\infty} q^{n-1} (\hat{\mu}(\alpha))^n = p \sum_{n=1}^{\infty} q^{n-1} = 1,$$

so that the Malthusian rate actually remains the same. But the other terms change drastically.

$$M_{\alpha}(dt) = e^{-\alpha t} M(dt) = p \sum_{n=1}^{\infty} q^{n-1} \mu_{\alpha}^{*n}(dt)$$

yields

$$\int_0^{\infty} t M_{\alpha}(dt) = p \sum_{n=1}^{\infty} q^{n-1} \int_0^{\infty} t \mu_{\alpha}^{*n}(dt) = p \sum_{n=1}^{\infty} q^{n-1} n\beta = \beta/p.$$

The classical variance decomposition applied to the sum of a geometrical number of independent random variables gives the variance

$$v^2/p + \beta^2 q/p^2.$$

Mutation probabilities are very small, so presumably this can be well approximated by just  $\beta^2/p^2$ , leading to the approximate time back to the first of  $n$  mutations as

$$Y + n\beta/p + \sqrt{n\beta/p} N(0, 1).$$

Here  $Y$ , having the stable age distribution of the macroprocess, is the only prob-

lematic part. With a probability  $r$ , the probability of non-extinction of the original process, it is exponential ( $a$ ), and with probability  $1 - r$  it has the stable age distribution of a process whose life length has the distribution of the conditional time to extinction of the original process, given that it dies out. Presumably it can be disregarded in a situation with little  $p$  and large  $n$ . We thus arrive at the final approximation

$$n\beta/p + \sqrt{n}\beta/pN(0, 1).$$

In molecular genetics what is known is however not  $n$  but the number  $d$  of differences between two existing species. If the same mutation rate is assumed for both species, as is usually done,  $n$  in the above formula should be replaced by a random variable that is binomial with parameters  $d$ ,  $1/2$ . This and a central limit theorem for sums of a random number of random variables leads to our final formula:

*The time back to separation between two species exhibiting at a certain amino acid chain  $d$  differences is, approximately,*

$$d\beta/2p + \sqrt{d/2} \sqrt{\beta^2/2p + v^2/p + \beta^2q/p^2}N(0, 1) \approx d\beta/2p + \sqrt{d/2} (\beta/2)N(0, 1)$$

**Example.** On a gene 80 differences have been found between two species. The time to their divergence has been estimated to 40 million years from other, *e.g.* fossile, evidence. If the generation time in the species considered is 10 years, equating  $d\beta/2p$  with 40 million yields  $p = 10^{-5}$ . If this were known to be the true probability of mutation a 95% "postdiction" interval for the time to divergence would be  $4 \times 10^7 \pm 1.3 \times 10^7$ .

Alternatively one could argue that fossiles only yield lower bounds for the time to appearance of a species. Then we would conclude that

$$40 \times 10/p + \sqrt{40} \times 10N(0, 1)/p > 4 \times 10^7,$$

and hence that

$$p < 1.36 \times 10^{-5}$$

with 99% confidence.

**4. Concluding Remarks.** This note sketches a branching process way of arguing about genetical evolution. It is very tentative. Serious application of this type of arguments would need more refined modelling. In particular the theory should incorporate multi-type individuals, so that mutation may possibly change the reproduction pattern. We hope to proceed to such work.



**References**

- [1] Barry, D. and Hartigan J. A. (1987) Statistical analysis of hominoid molecular evolution. *Statist. Sci.* **2**, 191-210.
- [2] Jagers, P. (1975) *Branching Processes with Biological Applications* . J. Wiley Sons, Chichester.
- [3] Jagers, P. (1989) General branching processes as Markov fields. *Stoch. Proc. Appl.* **32**.
- [4] Jagers, P. and Nerman, O. (1984) The growth and composition of branching populations. *Adv. Appl. Prob.* **16**, 221-259.
- [5] Kimura, M. (1979) The neutral theory of molecular evolution. *Sci. American* **241**, 94-104.
- [6] Nerman, O. (1984) *The Growth and Composition of Supercritical Branching Populations on General Type Spaces*. Dep. Mathematics, Chalmers U. Tech. and Gothenburg U. 1984-18.
- [7] Nerman, O. and Jagers, P. (1984) The stable doubly infinite pedigree process of supercritical branching populations. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **65**, 445-460.
- [8] Taib, Z. (1987) *Labelled Branching Processes with Applications to Neutral Evolution Theory*. Thesis, Department of Mathematics, Gothenburg U. To appear in Springer Lecture Notes in Biomathematics.
- [9] Wilson A. C. et al. (1987) Molecular time scale for evolution. *Trends in Genetics* **3**, 241-247.