Statistical Science
1991, Vol. 6, No. 3, 269–283

The Growth and Stabilization of Populations

Peter Jagers

Abstract. Population models in the birth-and-death style tradition have the unpleasant (and usually not advertised) implication that individuals do not age: It follows from the Markov properties of the whole population that life spans must be exponentially distributed and reproduction occur as splitting or in a Poisson process. This can be remedied only in parts (and at a high esthetical cost) by assuming more complicated Markovian properties in real time, like the age and parity dependent models of demography. Instead, if there is a sensible Markov structure in population growth, it resides in the pedigree, daughters inheriting genotypes from their mothers and being independent of their ancestors, given these types. This idea is used to define general branching processes and to analyze their properties: extinction, growth and asymptotic composition. The results are used to interpret the hypothesis of a molecular clock of mutations in biological evolution.

Key words and phrases: Branching processes, population dynamics, Markov fields, stable population theory, evolution, neutral mutations, molecular clock.

1. INTRODUCTION

Mathematical population theory is not the same as demography: Its object of study is not human populations. Nor is its object actual biological populations of, say, animals, bacteria or cells, or the physical populations of splitting particles in a cascade or neutron transport. Rather, its purpose is to study the common theme of these and many other empirical phenomena, an idealized pattern of free population growth, of sets changing as their members generate new set members.

The essence of such a theory is mathematical in the same sense as geometry, the study of idealized shape. It is relevant for actual populations in so far as their reproduction is close to the idealized free reproduction and to the extent that this reproduction property is important for the evolution of the system as a whole. Thus *in vitro* cell kinetics is close to the pattern, at least if the population has enough nutrition and space, whereas the well-regulated growth of a couple of fetus cells into, say, a hand is dominated by features other than population growth.

However, the population growth pattern is an important one, often playing a great role in the evolution of phenomena, and it can be discerned in many circumstances, ranging not only from demography to particle physics but including even data structures for sorting and searching in computer science (Aldous, 1991) or fractal sets arising in various types of mathematics (Larsson, 1990). Sometimes the conclusions you can draw from the general mathematical study are even stronger than those obtained through more specialized models. Many general truths have been rediscovered and launched as demographic, population dynamic or cell kinetic theories.

In this survey, we shall sketch the general Markov model of population growth, its structure, the growth and possible extinction of populations and the ultimate, stable composition of nonextinct populations. The latter will then be applied to interpret the famous hypothesis of a molecular clock of biological evolution.

2. THE STRUCTURE

Probabilistic population dynamics arises from the interplay of the population growth pattern with pure probability theory. (This is, of course, typical of much of so called "applied probability," not at all

Peter Jagers is Professor of Mathematical Statistics, Chalmers University of Technology and Gothenburg University, S-41296, Göteborg, Sweden. This article was a Special Invited Paper at the 213th IMS Meeting in Baltimore in 1990.

"applied" in the proper sense of the word but rather, like theoretical physics, arising from combining mathematics with idealized models of natural phenomena.)

Thus the classical Galton–Watson branching process couples the basic pattern of population growth with the very essence of classical probability, namely, addition of independent and identically distributed random variables; the population evolves from generation to generation by the individuals getting iid numbers of children. It has given rise to much intricate mathematics, maybe even too much, but also insight into at least one of the fundamental problems of actual populations, the *extinction problem* and its sequel, the question of possible size stabilization: If a freely reproducing population does not die out, what will happen to its size? Can it possibly stabilize, or must it grow beyond all bounds?

The tragic answer to these questions is no, there are no freely reproducing populations of stable sizes. And this is true for very general structures, much more general than the Galton–Watson process. Population size stability, if it exists in the real world, is the result of forces other than individual reproduction, of the interplay between populations and their environment.

There are many ways to see this mathematically, the basic reason being, of course, an underlying lack of balance between the properties of being large and very small, that is, extinct: However large a population might be, it can still die out but if it has died out it can never resurrect itself.

One of the many nice forms to make this precise can be found in Breiman's (1968) beautiful text book *Probability*, in the following exercise (page 98): Consider a sequence of nonnegative random variables $X_1, X_2, \ldots$ for which 0 is absorbing in the sense that $X_n = 0 \Rightarrow X_{n+1} = 0$. Assume that there is always a risk of extinction in the following way. For any x, there is a $\delta > 0$ such that

$$P(\exists\, n;\ X_n = 0 \mid X_1, \ldots, X_k) \geq \delta,$$

provided $X_k \leq x$. Then, with probability one, either there is an n such that all $X_k = 0$ for $k \geq n$ or $X_k \to \infty$ as $k \to \infty$.

The proof is quite direct from Lévy's theorem, $P(D \mid X_1, \ldots X_k) \to 1_D$ as $k \to \infty$, if D is measurable with respect to the σ-algebra generated by the X_i (as is $D = \{\exists\, n;\ X_n = 0\}$). This establishes the dichotomy between extinction or unlimited growth under extremely general circumstances.

But already the next natural question—what is the *rate* of this unlimited growth?—cannot be answered within the generation counting framework of Galton–Watson type processes. It requires the further structure of real time. We must know, not only how many children parents get, but also ages at child-bearings. Thus, the iid random variables describing reproduction have to be replaced by iid point processes, and the classical probabilistic addition of random variables by one of its outgrowths, the superposition of point processes. If this is performed according to the special pattern of population growth, then we can not only answer questions about rate of growth but also—and maybe this is more important—questions about the *ultimate composition* of nonextinct populations. What will the age distribution tend to be? What is the probability of being first-born? The average number of second cousins? Or the distribution of the time back to your nth grandmother's birth?

Many other composition questions cannot be posed, let alone answered, for the simple reason that the iid-ness we retain means that all individuals are considered to be of one and the same type. Thus, we are naturally brought on to *multi-type* branching populations: Whenever an individual is born, we know not only her mother's age but also her own *type*—think of it as a genotype, a starting platform from which the newborn's life is to evolve.

Mathematically, the individual reproduction process then turns into a point process on the product space, type × age. And the evolution of the newborn's life will no longer be decided in an iid fashion but rather according to a probability kernel, determined by the type of the newborn.

Thus, just as pure probability can be viewed as an edifice built on the straightforward analysis of sums of iid random variables, a successive hierarchy of applied probability models is obtained, by using more and more sophisticated probability theory, from addition of random variables to superposition of point processes, random measures and functions, and by moving away from independence to a Markov structure.

Formally, there is a *type space* $(S, \mathscr{S})$ and a *life space*, $(\Omega, \mathscr{A})$ of possible life careers, a probability kernel $P(s, \cdot)$, $s \in S$, on the life space, the *life law*, and finally a point process on $\mathrm{R}_+ \times S$, which is a random element defined on the life space, and suitably termed the *reproduction process*.

In this manner, the introduction of various types of individuals can be viewed as taking the step from independence to the simplest form of dependence in probability theory, Markovian dependence. You are born by your mother, who decides when you are to come into this world and also passes on a genotype to you. Given these two inherited properties, you lead your life independently of all your ancestors. This is the Markov model of population growth, the outcome of a straightforward combination of a vague population growth pattern with Markovianly dependent random lives and reproductions. And it leads to well-defined processes:

THEOREM 1 (Existence). *Each life law $P(s,\cdot)$, ancestral starting type $s \in S$ and the convention that the ancestor is born at time, say, $t = 0$, defines a unique general branching process with type space S.*

These processes are *branching*, meaning that, given their types, individuals who do not stem one from the other multiply independently of each other, that is, their futures and progeny are conditionally independent. They are also *Markov in the pedigree* that is, given her type, any individual lives and reproduces independently of all her ancestors. In other words, there is a conditional independence of the past given the present. Actually, the latter property implies the former:

THEOREM 2 (Markov is Branching). *If a population is Markov in the pedigree, then it must also be branching.*

A set of individuals is called a *line* (Neveu, 1986) if no member of it stems from another member. Thus, a typical line would be the kth generation, or part of it. But a line could also consist of chunks taken from different generations. It could even be random, the inclusion of members in it depending upon the process itself. An *optional* or *stopping* line is then a line such that the inclusion of an individual into the line is unaffected by her own and her progeny's lives. Like all other good Markov processes, branching is *strongly Markov*:

THEOREM 3 (Strong Markov). *Given the types and birth times of the elements of a stopping line, the populations stemming from the line elements constitute independent branching populations with the original life law, and the ancestral starting types and times given.*

For proofs, see Jagers (1989).

3. HOW TO MEASURE POPULATIONS

The general concept of additive measures of population size goes back almost two decades. With these, each individual is measured by a *random characteristic*, a stochastic process, whose value at time t is determined by:

- the individual's type,
- the individual's age now at time t,
- the individual's, and possibly all her progeny's life careers.

For simplicity, the characteristic is assumed to vanish for negative ages. In other words, individuals are not taken into account before they are born. The measure of the population at time t is the sum of all the characteristics, evaluated for all the individuals as above.

The simplest characteristic is, of course, the one that just records whether you are born or not, having the value of one if you are, and zero if you are not. Other characteristics may count the individuals below a certain age, of a certain type and so on and so forth (see Jagers, 1975).

Now fix some (decent) characteristic and, as time evolves, study the population as measured by that characteristic, a stochastic process in real time. Crucial for this is the *reproduction kernel* $\mu(s, A \times B)$, $A \in \mathscr{A}$, $B \in \mathscr{S}$, the number of B-type children that we should expect from an individual of type $s \in S$, while she is in the age-interval A. Under weak assumptions, it defines:

- a *Malthusian parameter*, α
- a *type reproductive value function*, h, and
- a *stable type distribution*, π.

(Mathematically, α is chosen so that the operator defined by the kernel $\int_0^\infty e^{-\alpha t}\mu(s,dt \times ds)$ has Perron root one; h and π are normed eigenfunctions and eigenmeasures of this operator. Biologically, α will turn out to be the exponential population growth rate, as hypothesized by Malthus, and h and π play the role indicated by their names. In the *one-type* case, $\#S = 1$, the reproduction kernel reduces to a measure, to be denoted by μ, on the nonnegative half-line, giving the expected number of children at various maternal ages. The Malthusian parameter is then the number rendering the Laplace transform of this measure one.)

To formulate the results we need notation for the Laplace transform of the characteristic, χ,

$$\hat{\chi}(\alpha) = \alpha \int_0^\infty e^{-\alpha t}\chi(s, \omega, t)\, dt,$$

where $s \in S$ stands for the type dependence, t for age and ω for the dependence on the lives of the individual and all her descendents. We denote its expectation for the case that the type s follows the stable type distribution, π, by $\mathbf{E}_\pi[\hat{\chi}(\alpha)]$.

THEOREM 4 (Expected Growth). *As $t \to \infty$, the expected size (with respect to a characteristic χ) of a population starting at time $t = 0$ from an ancestor of type s, behaves as*

$$h(s)e^{\alpha t}\mathbf{E}_\pi[\hat{\chi}(\alpha)]/\beta.$$

Here β is a constant (actually the long run average age at childbearing) and we have refrained from spelling out regularity conditions (see Jagers, 1989).

This result follows by rather typical applied probability methods, essentially Markov renewal theory. If it is conjoined with some classical pure probability, an intrinsic martingale, central limit theory, and some results on uniform integrability, it even follows that the process itself exhibits similar behavior:

THEOREM 5 (Population Growth). *Under slight regularity conditions, the population itself behaves as its expectation $h(s)$ replaced by a random variable w, which has $\mathbf{E}_s[w] = h(s)$ in the supercritical case, that is, $\alpha > 0$, is zero precisely when the population dies out, and is independent of the characteristic chosen.*

4. POPULATIONS IN VARYING ENVIRONMENT

From a Markovian point of view, the general branching population can be called time, or may be pedigree, homogeneous, since the life law remains the same whenever in time or wherever in the pedigree an individual is born. From a population dynamic aspect, this means that the environment does not change. In applications, environment may, of course, be highly varying, periodically with daylight, as for certain cell populations, or in more complicated ways. Thus, both mathematically and from the empirical side, it seems interesting to investigate situations where the life law exhibits a dependence on physical time.

Of course, in such a general context, there is faint hope for renewal type results about expected population size, although in some special cases, like the periodic, they can be derived (Jagers and Nerman, 1985). But if the process is normed by its expectation, so that the systematic part of time dependence is taken care of, then the same methods as in the time homogeneous case work out to show that the theorem on population growth still holds, of course under suitable regularity assumptions (Cohn and Jagers, 1991):

THEOREM 6. *Consider a population in varying environment, starting from an ancestor of type $s \in S$ at time $t = 0$. If population size at time t, normed by its expectation, is uniformly integrable over s and t, and the expected population size is "well-behaved" enough, then the normal population size converges in mean to a limiting random variable.*

5. THE ULTIMATE "STABLE" POPULATION

Even though much of what will be said can be extended to not-too-wildly varying environments, we turn back to supercritical branching populations in fixed environment. The special characteristic just counting all individuals born up to time t has the form $\chi(s, \omega, a) = 1$, if only $a \geq 0$, a being the age at t, and its Laplace transform has the value one. By the population growth theorems, the total population therefore grows like $we^{\alpha t}/\beta$. Dividing the general growth formula for an arbitrary (decent) characteristic by this special case, we conclude:

THEOREM 7. *As time passes, the average χ-value among all those born in a "decent" population not dying out converges to $\mathbf{E}_\pi[\hat{\chi}(\alpha)]$.*

In particular, this means that all proportions stabilize and hence the composition of the whole population has a limit, as time passes, the so-called *stable population composition*. To describe the latter explicitly is hard work, (see Nerman and Jagers, 1984) for the one-type case. (The multi-type case has not yet been published. It was, however, reported by Nerman, 1984.) But the theorem not only establishes the existence of the limiting composition, it also provides an explicit way to calculate various aspects of this stable composition from the reproduction kernel by choice of the appropriate characteristic. Thus, a bridge is established between the composition of populations and the probability laws of individual life.

I shall, however, venture to give a vague description of the limiting composition (for more details, see Jagers, 1991). This is most easily done by thinking of properties of a typical, that is, randomly sampled individual at a late time. Since sampling is performed among the total population of all those born, and this one increases exponentially at the Malthusian rate α, it is only natural that the typical individual's age should be, asymptotically, exponentially distributed with parameter α. As indicated by the name, her type will follow the stable type distribution π. The age will be independent of all other properties but the types backwards, that is, the type of the typical individual, her mother, grandmother and so on will constitute a Markov chain with transitions probabilities easily expressed in terms of the Malthusian parameter, the reproduction kernel and the stable distribution (Jagers, 1991). And the times backwards to the successive births interact with the types in a Markov renewal manner.

Forwards properties of the typical individual herself and her descendents will be as in the original branching process.

In one-type populations (where the reproduction kernel reduces to a reproduction measure, μ, and h and π are trivially one), the stable composition is much more easily described, since there is no

Markovian wandering between types, as we look backwards into the past. Thus, if the typical individual's age is called X, her mother's age at giving birth to her Y_1, grandmother's age at begetting mother Y_2, and so on, then X and all the Y_i will be independent, the latter iid with the distribution function

$$\int_0^t e^{-\alpha u}\mu(du),$$

known as the *stable age at childbearing*. Its expectation is

$$\beta = \int_0^\infty u e^{-\alpha u}\mu(du),$$

the mean age at childbearing, known from the growth theorems.

This means that the successive birth times backwards constitute a delayed renewal process, and by the renewal theorem the expected number of these births during time t back is roughly t/β, at $t \to \infty$.

Now, you may get the impression that this is all very theoretical and typical of mathematicians to be interested in proportions of the *total* population when it is the *living* population that really counts. However, the step from one to the other is very simple: just condition on being alive. For example, if the life span distribution is L, in the one-type case, then, since the typical individual in the total population leads her life independently of her age at sampling, the age distribution in the living population is directly calculated to be

$$\frac{\int_0^t e^{-\alpha u}(1 - L(u))\,du}{\int_0^\infty e^{-\alpha u}(1 - L(u))\,du},$$

the classical stable age distribution.

6. THE MOLECULAR CLOCK OF EVOLUTION

Biological evolution provides a field where population growth has been going on for a very long time, and interest is focused upon the development (through mutations) up to the present situation. This is an appealing situation for analysis in terms of the backwards structure of a stabilized branching population.

The much-discussed neutral theory of evolution claims that much of evolution is due to selectively neutral mutations (Kimura, 1983). The idea of a mutational clock of evolution adds to this that the mutations, in a particular gene or protein, like the much studied globins, should have occurred in a Poisson-like stream, with an intensity that is typical of the gene, rather than the species, its reproduction pattern or population size. Neutral mutations, by definition, do not change (evolutionarily relevant aspects of) the reproductive pattern of individuals. Therefore they are suitably studied in terms of branching processes, and at a first approximation even one-type branching processes. To model them, introduce a probability of mutation into one-type branching populations. Let this probability possibly be influenced by the mother's age at giving birth to the mutant or nonmutant child, and denote it by $p(u)$ at the maternal age u. Following Taib (1987), write

$$\mu_m(t) = \int_0^t p(u)\mu(du)$$

for the expected number of mutant children up to age t. (On the surface more generally, you may start from μ_m, mutations possibly being influenced by many things other than age, and then use the Radon-Nikodym theorem to get the above form.) Now merge a mutant individual with all her nonmutant progeny into what Taib calls a *macro-individual*. Obviously, these constitute a new branching population.

The reproduction function, that is, kernel in the one-type case, turns out to be

$$M = \sum_{n=0}^{\infty} (\mu - \mu_m)^{*n} * \mu_m,$$

corresponding to the various generations where the first mutant descendant might appear. (The asterisk denotes convolution.) It may seem astonishing, but it is easy to check that this measure has the same Malthusian parameter α as the original reproduction function,

$$\hat{M}(\alpha) = \sum_{n=0}^{\infty} \left(\hat{\mu}(\alpha) - \hat{\mu}_m(\alpha)\right)^n \hat{\mu}_m(\alpha)$$
$$= \sum_{n=0}^{\infty} \left(1 - \hat{\mu}_m(\alpha)\right)^n \hat{\mu}_m(\alpha) = 1.$$

Hence the expected age at childbearing in the macropopulation is easily calculated to be

$$\int_0^\infty t e^{-\alpha t} M(dt) = \beta/\hat{\mu}_m(\alpha).$$

Thus, by our earlier results the process of macrobirths backwards is a renewal process with intensity $\hat{\mu}_m(\alpha)/\beta$. But this is, of course, precisely the process of mutations up to the present situation. For the process to have the same asymptotic rate for all species with different reproduction patterns, there must thus be a constant c, the *evolution rate*, such that

$$\hat{\mu}_m(\alpha)/\beta = c$$

For all possible μ. It is again easy to check that this is the case if and only if the mutation probability is linear in the mother's age, $p(u) = \lambda u$, that is,

λ is the *mutation rate* and that then $c = \lambda$. We summarize:

THEOREM 8 (Molecular Clock of Evolution). *The process of mutations up to the present stage is a renewal process for any species. It is always independent of population size. Its rate, the evolution rate, is the same for all species with the same mutation rate, and the two rates are equal.*

COROLLARY 1. *Assume that mutation is the result of at least one of a Poisson stream of genetic events with intensity* λ. *If this is small, then the resulting mutation probability at maternal age* u, $p(u) = 1 - e^{-\lambda u}$, *yields an approximately species-independent molecular clock of mutations. The evolution rate is approximately the mutation rate.*

These results thus hold in a one-type branching process model of evolution. Such a model is obviously simplistic and should be looked upon as first try, indicating what type of results one could obtain in more general studies. Still it has certain advantages above traditional population genetic models:

- Population size need not be fixed.
- We are not restricted to discrete time models with nonoverlapping generations or to simple birth-and-death style processes with exponentially distributed life spans.
- Completely general patterns of reproduction can be allowed.

The drawback is, of course, the asexual character of branching processes. But that might be of less importance in the present context than in the type of applications many population genetic models have been devised for, situations like breeding, where mating is a fundamental property.

ACKNOWLEDGMENT

This article was supported by a grant from the Swedish Natural Science Research Council.

REFERENCES

ALDOUS, D. (1991). Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1** 228–266.

BRIEMAN, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

COHN, H. and JAGERS, P. (1991). *General Branching Processes in Varying Environment*. To appear.

JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley, Chichester.

JAGERS, P. (1989). General branching processes as Markov fields. *Stochastic Process. Appl.* **32** 183–212.

JAGERS, P. (1991). Stabilities and instabilities in population growth. Unpublished manuscript.

JAGERS, P. and NERMAN, O. (1984). The growth and composition of branching populations. *Adv. in Appl. Probab.* **16** 221–259.

JAGERS, P. and NERMAN, O. (1985). Branching processes in periodically varying environment. *Ann. Probab.* **13** 254–268.

KIMURA, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press.

LARSSON, P. (1990). L'ensemble différence de deux ensembles de Cantor aléatoires. *C. R. Acad. Sci. Paris Ser. I Math.* **319** 10, 735–738.

NERMAN, O. (1984). The growth and composition of supercritical branching populations on general type spaces. Technical Report, Dept. Mathematics, Chalmers Univ. of Technology and Gothenburg Univ. 1984-18.

NERMAN, O. and JAGERS, P. (1984). The stable doubly infinite pedigree process of supercritical branching populations. *Z. Wahrsch. Verw. Gebiete* **65** 445–460.

NEVEU, J. (1968). Arbres et processus de Galton-Watson. *Ann. Inst. H. Poincaré Probab. Statist.* **22** 199–207.

TAIB, Z. (1987). Labeled branching processes with applications to neutral evolution theory. Ph.D. dissertation, Dept. Mathematics, Chalmers Univ. of Technology and Gothenburg University. (To appear as a Springer Lecture Note in Biomathematics.)