Stochasticity in the adaptive dynamics of evolution: the bare bones

Fima C. Klebanera, Serik Sagitovb, Vladimir A. Vatutinc, Patsy Haccoudd and Peter Jagersb*

aDepartment of Mathematical Sciences, Monash University, Melbourne, Australia; bDepartment of Mathematical Sciences, Chalmers University of Technology, Gothenburg, Sweden; cSteklov Mathematical Institute Moscow, Russian Academy of Science, Moscow, Russia; dInstitute of Environmental Sciences (CML) and Mathematical Institute, Leiden University, Leiden, The Netherlands

(Received 25 October 2009; final version received 23 June 2010)

First a population model with one single type of individuals is considered. Individuals reproduce asexually by splitting into two, with a population-size-dependent probability. Population extinction, growth and persistence are studied. Subsequently the results are extended to such a population with two competing morphs and are applied to a simple model, where morphs arise through mutation. The movement in the trait space of a monomorphic population and its possible branching into polymorphism are discussed. This is a first report. It purports to display the basic conceptual structure of a simple exact probabilistic formulation of adaptive dynamics.

Keywords: adaptive dynamics; branching process; speciation

1. Driving simplicity to its extreme

One aspect of the fascination with biology is the contrast between the overwhelming detail of nature and its order. Deus creavit, Linnaeus disposuit, was Linnaeus’ famous, not so humble summary of the two sides and who cared for which. The friendly and modest country gentleman Darwin had little of Linnaeus’s self-image, but shared his obsession with the nitty-gritty of flora and fauna, since early childhood. The idea of evolution must have come forth as gratifyingly simple to him and to any biologist immersed in the richness of nature. To a mathematician, it invites further reduction into a more or less axiomatic theory, where everything can be derived from first principles.

Adaptive dynamics is an attempt at this. In many regards, it is successful albeit (of course!) criticized for simplistic genetics as well as ecology. A critique more on grounds of principle is that it treats populations as infinite and even as a continuous matter. The renowned canonical equation [5] is an example of this.

Recently, Méleard, Champagnat, Lambert and co-authors have published a sequel of seminal articles [2–4], and others, reformulating classical adaptive dynamics [7,20,21] in terms of...
population-size-dependent birth-and-death processes, then passing to various limits by letting first the carrying capacity grow, so that populations become infinite, then mutational steps shrink to zero. The present paper presents another such oversimplified model, going back to the ideas of [1,26,27], and indeed in a certain sense the simplest possible, still retaining indispensable properties of a population under evolution. In our view, they are as follows.

1. Any such population is finite and consists of reproducing individuals.
2. Individuals are characterized by their traits.
3. Together with population size (of the own species as well as competitors) traits may influence reproduction.
4. But they do not determine it: there is a variation in the reproductive behaviour between individuals of the same kind and, seemingly, in the same circumstances.
5. At reproduction, there is a probability for mutation, resulting in a child with a trait different from its mother.
6. The resulting population survives or not, depending upon chance and fitness.

The finiteness requirement is essential: natural populations are finite. Still, passage to limits of infinite populations may help to uncover underlying structure, or exhibit its aspects more clearly. But this is a tool only and it should occur explicitly, so as not to hide difficulties, avoid mistakes and also render investigations into the accuracy of approximations feasible. There is no all-embracing continuity of properties from the finite to the infinite; infinite populations need not die out, even though their finite counterparts necessarily do, if bounded and allowing variation in reproduction between individuals. Usually, in models, furthermore, not only is the carrying capacity of the habitat large (tending to infinity) but also mutation steps small (infinitesimal), and two limits thus involved, the order between them possibly significant.

Variation is another intrinsic feature of nature. A deterministic description of individual life remains an illusion: there is randomness in life. The simplest stochastic reproduction conceivable assumed here is asexual binary splitting. Following the adaptive dynamics tradition, we disregard mating. This will lead to underestimation of extinction time, but simplifies theory radically. Thus, we assume that each individual either gets two children in the next generation or none.

Similarly, interaction with environment and population is drastically condensed. The habitat concerned has a carrying capacity $K$, and if there is only one population present, of size $n$, the probability of an individual successfully splitting is taken to be $K/(K+n)$. Given the population size, individuals reproduce independently. If there are several morphs present, the number $n$ is replaced by a linear combination of the various population numbers and coefficients expressing the strength of competition. Again, a Gaussian form of these coefficients will be assumed in the last section. But first we consider just one or two morphs, with given coefficients.

The carrying capacity can be different for different species; $K$ can depend upon the trait considered. Then we write $K(t)$, $t$ being a real number, the trait, which can be thought of as abstract, like a reproductivity measure, or in a very concrete manner, like the length of a beak. In the last section, we shall work with a specific Gaussian form of dependence of carrying capacity upon trait.

At reproduction, mutation may occur with some small probability which we will assume to be neither trait nor population-dependent; with a probability $\mu$, a newborn with a mother with trait $t$ turns out to be a mutant, then with a random trait normally distributed around $t$.

Time structure is stripped to its bare bones: only the successive (non-overlapping!) generation size are counted, or alternatively everybody is supposed to have the same life-span of one season.

Clearly, when compared with the many-faceted phenomenon of evolution, such a mathematical sketch is nothing but a caricature. Still, though reproduction and time structure are the simplest possible, they retain a fundamental affinity to reality. Other properties are more arbitrary, like the
form of dependence on the environment or the normal distribution of mutational steps or in the form of certain coefficients. The former has some historical precedence, in the papers cited and earlier, but nothing else to speak for it than its ability to yield a reproductive behaviour changing with the relation of population size to carrying capacity and prevalence of competing species. The assumption of normal distributions is also *ad hoc* though a central limit style argument from many small effects might have passed through the minds of the first proponents of this distribution for mutational steps. (In other contexts, it has been argued that mutational steps should follow extreme value distributions, since only the most advantageous mutations result in viable children [23].) But actually, we shall not use more of the normal distribution of mutational steps than the fact that deviations are of the order of the standard deviation, a fact that is much more generally true. And in many regards, this specific model is just a representative, and many of its properties, like those building on second moment properties, large deviation theory, quasi-stationarity and early growth and ultimate extinction, extend to more general structures.

In the next two sections, neither the mutation probability nor any mutation step distribution will occur, though in Section 3 the fact that we consider two populations, one of which called the *resident* starts from its carrying capacity, the other called the *mutant* population starting from one individual, certainly reflects having a mutation probability in mind, which is taken to be small enough for the first population to have grown to its carrying capacity before the mutation occurs. In Section 4, these mechanisms will be made explicit. The difference from the series of papers by Champagnat, Ferrière, Méléard and Lambert lies in our taking discrete time, non-overlapping generations, i.e. Galton–Watson-style processes as the simple ‘pure’ structure for first study.

### 2. A single trait

The model population is, thus, a binary, population-size-dependent Galton–Watson branching process. It starts from a positive integer number $Z_0 = z$ and is then recursively given by

$$ Z_{n+1} = \sum_{j=1}^{Z_n} \xi_{nj}, $$

$$ P(\xi_{nj} = 2|Z_n) = \frac{K}{K + Z_n}, \quad P(\xi_{nj} = 0|Z_n) = \frac{Z_n}{K + Z_n}. \quad (1) $$

The random variable $\xi_{nj}$ is the number of children of individual $i$ in generation $n$. These are independent and identically distributed, given the population size $Z_n$, or indeed the whole past population history, $Z_0, Z_1, \ldots, Z_n$. We refer to the process value $Z_n$ as *population size* (at time or generation $n$). Since reproduction is identically distributed for all individuals in the same generation and the distribution, given $Z_n$, is the same for all generations $n$, we shall often delete the suffices, at least when not referring to several individuals in one context.

The trait dependence of the carrying capacity $K$ is suppressed, since we consider only one trait, and no mutations yet. $K$ should be thought of as a large number, and later we shall let it tend to infinity.

It is important to grasp this single trait process not only for its own sake but also for the analysis of the complete model with mutation and several traits.

Obviously, it behaves like a subcritical branching process whenever the population is greater than $K$, like a supercritical process for all population sizes smaller than $K$, and it is critical if the size is precisely $K$ (then necessarily an even integer).

Since thus the population has a bounded expectation, it follows that the extinction probability is $P(Z_n \to 0, \text{ as } n \to \infty) = 1$ (cf. [8, pp. 108 ff.] or [13]). Ultimately, it dies out.
Under the level $K$, the population is, however, supercritical. Therefore, it tends to increase, while this is the case, and it is prone to reach large values (around $K$) before ultimate extinction. In fact, we shall show that the time to extinction is very large, with an overwhelming probability. The population will settle at a quasi-stationary regime, in the sense that it will fluctuate around the value $K$ for a time period that is exponentially long in $K$, until it finally dies out.

For $a \geq 0$, write $T_a$ for the first time the population reaches, or passes, $a$, from below or above, depending upon the starting position. For short, let $T = T_0$ be the time of extinction. What is the relation between these two random variables for large $a$? Indeed, it is only rarely that the process dies out before reaching a level $dK$, if $0 < d < 1$ is small or the starting number $z$ large.

**Proposition 1** Let $0 < d < 1$. Then for any $1 \leq z \leq dK$,

$$P_z(T < T_{dK}) < d^z.$$

In this, and elsewhere, the probability or expectation indexed by $z$, $P_z$, $E_z$ means the probabilities of such a population starting from size $Z_0 = z$. No index either indicates $z = 1$, or else some completely different probability measure.

**Proof** Such assertions are proved by comparison with suitably chosen (not population size dependent) simple Galton–Watson branching processes, about which much is known. In the present case, consider such a binary splitting process with the probability of begetting zero children being $d/(d + 1)$. Call it $\hat{Z}_n$. Since $x/(1 + x)$ is an increasing function of $x$, any $k < dK$ yields

$$\frac{k}{K + k} < \frac{dK}{K + dK} = \frac{d}{d + 1}.$$

Hence, as long as population size stays below $dK$, the probability of producing no offspring is smaller than the corresponding probability pertaining to this classical Galton–Watson process. Therefore, clearly the probability that our process becomes extinct by time $n$, without crossing $dK$, is smaller than the corresponding probability for the binary Galton–Watson process $\hat{Z}_n$. The latter must be smaller than the Galton–Watson probability of ultimate extinction, $P(\hat{Z}_n \rightarrow 0) = \hat{q}$. It is well known (cf. [8, p. 113]) that $\hat{q}$ is the smallest root of the quadratic equation

$$\frac{d}{d + 1} + \frac{1}{d + 1}\hat{q}^2 = \hat{q},$$

which is simply $d$. Hence,

$$P_z(T < T_{dK}) = P_z(Z_n = 0 \text{ for some } n < T_{dK}) \leq P_z(\hat{Z}_n \rightarrow 0) = d^z.$$

Thus, with a positive probability, the population will not die out but reach sizes at the order of the carrying capacity. Sooner or later, it will then die out. But when?

**Proposition 2** Whatever the starting number $z$, carrying capacity $K$ and time (generation) $n$,

$$P_z(T > n) \leq (1 - e^{-K})^n \leq \exp[-ne^{-K}]$$

and $E_z[T] \leq e^K$. 


Proof. Write

\[ Q_n := P_\zeta(T > n). \]

The elementary inequality

\[ \left( \frac{k}{K + k} \right)^k \geq \left( \frac{1}{1 + Kk^{-1}} \right)^k \geq e^{-K}. \]

yields

\[ P_\zeta(T \leq n + 1) = P_\zeta(T \leq n) + \sum_{k=1}^{\infty} P_\zeta(Z_n = k) \left( \frac{k}{K + k} \right)^k \]

\[ \geq P_\zeta(T \leq n) + e^{-K}Q_n. \]

Hence,

\[ Q_{n+1} \leq Q_n - e^{-K}Q_n, \]

and the upper bounds on the probabilities follow by induction and another elementary inequality, \(0 < 1 - u < e^{-u}\) for \(0 < u < 1\). The second assertion follows by summation:

\[ E\zeta[T] = \sum_n P_\zeta(T > n) \leq \sum_n (1 - e^{-K})^n = e^K. \]

A corresponding lower estimate can be deduced but it is irrelevantly small. Instead, note that we already know that with a positive probability the process will move from any starting point up to the vicinity of the carrying capacity. By large deviation theory, it will stay there a very long time until by a random perturbation it leaves the area (if it were deterministic, it would get stuck) and by chance becomes extinct. This is why the upper exponential bound derived indeed gives a survival time of the right order; persistence time is exponential in the carrying capacity. We shall have a closer look at this now.

Introduce the density process \(X_n = Z_n/K\), and note that the offspring distributions are in fact functions of the density,

\[ P(\xi = 2 | Z_n) = \frac{1}{1 + X_n}, \quad P(\xi = 0 | Z_n) = \frac{X_n}{1 + X_n}. \]

We have that

\[ X_{n+1} = \frac{1}{K} \sum_{j=1}^{KX_n} \xi_{nj}, \]

where the \(\xi_{nj}\) are the offspring numbers in the \(n\)th generation, explicitly indexed, and are indeed independent and distributed like twice a Bernoulli random variable with the parameter \(p(X_n) = 1/(1 + X_n)\). Writing \(m(X_n) = 2p(X_n)\) for the offspring mean when the density is \(X_n = Z_n/K\), and centring the variables of the sum in the usual central limit fashion, we obtain

\[ X_{n+1} = \frac{1}{K} \sum_{j=1}^{KX_n} (\xi_{nj} - m(X_n)) + X_n m(X_n). \]

With

\[ f(x) = xm(x) = \frac{2x}{1+x}, \]
and
\[ \eta_{n+1} = \frac{1}{\sqrt{K}} \sum_{j=1}^{K X_n} (\xi_{nj} - m(X_n)), \]
we can write this
\[ X_{n+1} = f(X_n) + \frac{1}{\sqrt{K}} \eta_{n+1}. \] (2)

Clearly, \( E[\eta_{n+1}] = 0 \) and
\[ E[\eta_{n+1}^2 | X_n] = X_n \ Var(\xi | X_n) = \frac{4X_n^2}{(1 + X_n)^2}. \]

Hence, a central limit theorem argument shows that the noise term in Equation (2) is of order \( 1/\sqrt{K} \).

It follows that the process \( X_n \), which does of course depend upon \( K \), converges (in various senses), as \( K \to \infty \), to the deterministic sequence \( \{x_n\} \), defined by an initial value \( x_0 \) and the recurrence \( x_{n+1} = f(x_n) \) [14]. This is a sort of law of large numbers; a corresponding central limit theorem is available in [16].

However, this convergence also pinpoints the difference between stochastic and deterministic systems. Indeed, if \( x_0 = 0 \), then for all \( n \), \( x_n = 0 \) since 0 is a fixed point of \( f \). In its awkward way, this is how the deterministic system tells us that extinction is certain. For other outcomes, we need \( Z_0/K \to x_0 > 0 \), as \( K \to \infty \). In other words, the population should become comparable to \( K \) in order to get attracted to the stable fixed point 1. However, by Proposition 1, we know that this will occur with a high probability, and after that \( X_n \) will be approximated, for large \( K \), by \( x_n \) with a positive starting value. The process will be near 1 for late times \( n \) and this means that the original population size process \( Z_n \) will be near \( K \). The deterministic dynamics being globally attractive, \( Z_n \) will aim at moving towards \( K \); only the event of extinction can prevent this.

Now, due to the random nature of the density process, exhibiting small but persistent perturbations from \( x_n \), ultimately there will be an exit from the levels \([dK, d'K]\), \( 0 < d < 1 < d' \), due to large deviations of random terms.

The main theory for such systems is analogous to the Freidlin–Wentzell theory and was developed in [12], but the present special setting, the problem of exit from a domain, was solved in [17]. Under some assumptions, it was shown that if the deterministic system has a stable fixed point, then the perturbed system will stay in the domain of attraction of this point for an exponentially long time, \( e^{cK} \) for some \( c > 0 \). Since in our case the domain of attraction is the whole positive line (but not zero), this result applies to the extinction time. For any \( \delta > 0 \) and a starting point \( z \geq dK \),
\[ P_z(\xi(e^{(c-\delta)K} < T < e^{(c+\delta)K}) \to 1, \quad K \to \infty. \]

In the special case of our model, where summands are (twice) binomial random variables, there is a direct proof, using a large deviation result due to Janson [11], stating \( c \) explicitly. Indeed,
\[ c = \frac{d(1 - d^2)}{8(1 + d)}, \] (3)
see also [9]. Since, in the present context, our interest in this process is rather as a representative of a broad class of population models, proof and results specific for that model will be published elsewhere [10].

If the process starts from a small number \( z \) of individuals, then the time it takes to reach the level \( dK \) (if it does not die out) is at least of an order approximated by the binary Galton–Watson process with mean \( 2/(1 + d) \) and at most of an order corresponding to such a process with mean
reproduction $2K/(K+1)$, i.e. of order log $dK$, since for any Galton–Watson process $\hat{Z}_n$ we will have

$$\hat{Z}_{T_{dK}} \approx dK.$$ 

By exponential growth, the conclusion will follow that the time to reach $dK$ is of the order log $dK$; for details cf. [10].

Close to $K$, our process is near critical and approach will be slower; it is well known that a non-extinct critical branching process tends to grow linearly [8], and our process being on the supercritical side will tend to reach the carrying capacity quicker than in linear time $(1-d)K$. Actually, it can be proved that linear time is correct [10]. In conclusion, if the population does not die out before reaching high levels, the time until extinction is of the order $e^{cK}$, only a minor part (of the order $K$) is spent growing up to a quasi-stationary size. That one also has a limiting distribution.

**Proposition 3** Consider the distribution of $X_n$, given that $X_n > 0$, for fixed $K$. As $n \to \infty$, this converges to a proper distribution function called the quasi-stationary distribution. When in this setting $K \to \infty$, all probability concentrates at the point 1.

This was shown in [15] and is also contained in the more general setting of [18,25]. Indeed, the existence of quasi-stationary distributions is a consequence of the Krein–Rutman theory of positive operators.

Simulations exhibit first the quick growth and ensuing quasi-stabilization of five populations starting from one ancestor (Figure 1) and then the pseudo-stable distribution (Figure 2).

### 3. Two morphs

After some, probably a long, time, the population will experience its first mutation giving rise to a new population. By then, the original resident population will be in its quasi-stationary stage; mutation probabilities are thus taken to be small. The new, mutant population has just one member, its ancestor. The basis of adaptive dynamics can then be said to be furnished by the branching process fact that the new population either dies out or else embarks on exponential growth, in which case the old resident dies out, or the two populations will coexist for a time span that is exponential in the carrying capacity. We proceed to make this precise in terms of our model.

The process starts from a pair $(Z_{0}^{(1)}, Z_{0}^{(2)})$ of positive integers, the first component denoting the size of the resident and the second that of the mutant population, when the mutation appears.
Thus, $Z_0^{(1)}$ is in the vicinity of the carrying capacity and $Z_0^{(2)} = 1$. Transitions from generation $n$ to $n + 1$ are described by

$$Z_{n+1} = (Z_{n+1}^{(1)}, Z_{n+1}^{(2)}) = \left( \sum_{k=1}^{\xi_{nk}^{(1)}} Z_n^{(1)}, \sum_{k=1}^{\xi_{nk}^{(2)}} Z_n^{(2)} \right),$$

where $\xi_{nk}^{(i)}, k = 1, 2, \ldots$, are independent given the preceding generation $Z_n$ and satisfy

$$P(\xi_{nk}^{(1)} = 0|Z_n) = \frac{Z_n^{(1)} + \gamma Z_n^{(2)}}{a_1 K + Z_n^{(1)} + \gamma Z_n^{(2)}},$$

$$P(\xi_{nk}^{(1)} = 2|Z_n) = \frac{a_1 K}{a_1 K + Z_n^{(1)} + \gamma Z_n^{(2)}},$$

and

$$P(\xi_{nk}^{(2)} = 0|Z_n) = \frac{\gamma Z_n^{(1)} + Z_n^{(2)}}{a_2 K + \gamma Z_n^{(1)} + Z_n^{(2)}},$$

$$P(\xi_{nk}^{(2)} = 2|Z_n) = \frac{a_2 K}{a_2 K + \gamma Z_n^{(1)} + Z_n^{(2)}}.$$

Now, this process has two competing types, but it is not genuinely multi-type, in the sense that none of them can produce children with the other trait. We also disregard, presently, further mutation to new traits. The interaction coefficient $\gamma$ is supposed to be positive but less than one, and the same in both directions. This means that interaction is taken as symmetric, an assumption that is not at all necessary. Indeed, taking them different can lead to several interesting phenomena, like that of evolutionary suicide [9,24]. It may be noted that though the specific form of dependence is certainly \textit{ad hoc}, it has some historical precedence, going back at least to [19].

We assume that the resident population starts exactly at its carrying capacity, which is $a_1 K$. Then the mean reproduction of the mutant in a generation with $z$ mutants present will be

$$\frac{2a_2 K}{a_2 K + \gamma a_1 K + z} > 1 \iff z < (a_2 - \gamma a_1) K.$$

It follows that the mutant is subcritical throughout, if $a_2 < \gamma a_1$, and indeed dominated by a subcritical binary Galton–Watson process with the splitting probability $a_2/(a_2 + \gamma a_1)$. By [22],

![](figure2.png)

Figure 2. Histogram of a population size for the last 500 of 10,000 generations with $K = 50$. 
the probability that the maximum $M$ of the latter reaches $j$ is of the order

$$P_1(M \geq j) \sim \frac{(a_2/\gamma a_1)^j}{j}.$$ 

Hence, for large carrying capacities, the mutant population will die out directly, never coming close to $K$ with overwhelming probability.

If instead

$$a_2 > \gamma a_1,$$  \hspace{1cm} (4)

and $K$ large enough, the mutant population will instead start supercritically, and either die out as above or else grow at a geometric rate for some time, much as in the preceding section. This is the case of possible invasion. It is illustrated in the following figure, which shows five runs of a population system with $\gamma = 0.7, a_1 K = 40$ and $a_2 K = 70$. In three cases the resident prevailed, the invader dying out very quickly, and in two runs, the invader took over (Figure 3).

We proceed to have a look at the qualitative behaviour of the process and the domains of super- and subcriticality (Figure 4). They will be given in terms of the $(x_1 = z_1/K, x_2 = z_2/K)$-plane in which the process has a supercritical, critical and subcritical reproduction, respectively. Solving the inequalities for the mean offspring number greater than one yields that these regions are

Figure 3. Five competitive population evolutions, $K = 100, \gamma = 0.7, a_1 = 0.4$ and $a_2 = 0.7$. In two of them (red and orange lines), the invader takes over.

Figure 4. Regions of super- and subcriticality.
defined by:

\[ x_1 + \gamma x_2 < a_1, \quad \gamma x_1 + x_2 < a_2, \]

\[ x_1 + \gamma x_2 = a_1, \quad \gamma x_1 + x_2 = a_2, \]

\[ x_1 + \gamma x_2 > a_1, \quad \gamma x_1 + x_2 > a_2. \]

Note that the boundary is where one of the populations is critical; it consists of two segments of straight lines. The point at which both processes are critical is

\[ x_1^* = \frac{a_1 - \gamma a_2}{1 - \gamma^2}, \quad x_2^* = \frac{a_2 - \gamma a_1}{1 - \gamma^2}. \]

Since \( 0 < \gamma < 1 \), there are three possibilities for this point

\[ x_1^* > 0, \ x_2^* > 0; \quad x_1^* < 0, \ x_2^* > 0; \quad x_1^* > 0, \ x_2^* < 0. \]

The remaining case \( x_1^* < 0, x_2^* < 0 \) can be ruled out, since it implies both that \( \gamma > a_1/a_2 \) and that \( \gamma > a_2/a_1 \), and hence \( \gamma > 1 \).

As an example, we give the following picture for \( x_1^* > 0, x_2^* > 0 \) made with \( a_1 = 0.8, \ a_2 = 0.99, \ \gamma = 0.7 \) and where we take \( K \) as the scale unit of the graph.

The domain A above the two lines corresponds to subcriticality of both populations, the left intermediate domain B is where the mutant is subcritical and the resident supercritical and the right intermediate domain C corresponds to supercriticality of \( \{ Z^{(2)}_n \} \) and subcriticality of \( \{ Z^{(1)}_n \} \), whereas the domain below the two lines corresponds to their being supercritical. This is precisely the situation where coexistence may be realized.

Like in the single trait setup, the process \( \{(Z^{(1)}_n, Z^{(2)}_n)\} \) scaled by \( K \), the density process, can be approximated by a two-dimensional dynamical system for large values of \( K \). The latter has a fixed point at \( (x_1^*, x_2^*) \) and is attracted to it. Combine this with the fact that the density process is non-negative to see that the coexistence case is when the fixed point lies in the positive quadrant \( x_1^* > 0, x_2^* > 0 \) and corresponds to the condition

\[ \gamma < \frac{a_1}{a_2}, \quad \gamma < \frac{a_2}{a_1}. \]

Like in the single trait case, the two-trait process must die out. But the process being supercritical for small population sizes, both populations will settle to a quasi-stationary regime, in the sense that their sizes will fluctuate around the values \( Kx_1^* \) and \( Kx_2^* \) for an exponentially long time (in \( K \)), provided the new trait gets established, in the sense that its size becomes comparable to \( K \).

Thus, the probability of this happening is needed. Estimates from below and from above the probability of more or less direct extinction are established in the following proposition, pointing at \( \gamma a_1/a_2 \) as a rough indicator. Superscripts refer to the corresponding populations. We omit the suffixing starting point (which is \( (a_1 K, 1) \)).

**Proposition 4** The probability of the mutant dying out while the resident remains above the level \( d K, 0 < d < a_1 \), is at least

\[ P(T^{(2)} < T^{(1)}_{dK}) > \frac{\gamma d}{a_2}. \]
Similarly, the probability of the mutant population dying out before it reaches the level $bK$, $b < a_2$, while the resident remains below the level $uK$, $u > a_1$, is

$$P(T(2) < T^{(1)}_{bK} \land T^{(1)}_{uK}) < \frac{\gamma u + b}{a_2}. \quad (7)$$

In this $b, u$ can be so chosen that $\gamma u + b < a_2$.

**Proof** On the set $\{T(2) < T^{(1)}_{uK}\} \cap \{T(2) < T^{(2)}_{bK}\} \cap \{n < T(2)\}$, $Z^{(1)}_n < uK$ and $Z^{(2)}_n < bK$. Hence, the probability of the new trait to reproduce successfully is

$$P(\xi_k = 2 | (Z^{(1)}_n, Z^{(2)}_n)) = \frac{a_2 K}{a_2 K + \gamma Z^{(1)}_n + Z^{(2)}_n} > \frac{a_2}{a_2 + \gamma u + b}.$$ 

Provided $\gamma u + b < a_2$, the splitting probability is larger than one-half. Selecting $u$ in the range $a_1 < u < a_2/\gamma$ allows us to choose $0 < b < a_2 - \gamma u < 1$. Consequently the trait extinction probability does not exceed that of a supercritical Galton–Watson process with the splitting probability as above. This leads to the upper bound. Bounding the splitting probability from above on the set $\{T(2) < T^{(1)}_{dK}\}$ yields

$$P(\xi_k = 2 | (Z^{(1)}_n, Z^{(2)}_n)) = \frac{a_2 K}{a_2 K + \gamma Z^{(1)}_n + Z^{(2)}_n} < \frac{a_2}{a_2 + \gamma d}$$

and the lower bound for the extinction probability of the mutant. \hfill \blacksquare

We proceed to bounding the probability of the mutant dying out, while the resident remains close to the carrying capacity.

**Proposition 5** Let $Z_0 = (z, 1)$. For $z > dK$, the mutant survival probability

$$r_{n,n} := P(T(2) > n | T^{(1)}_{dK} > n)$$

satisfies

$$r_{n,n} \geq \left(1 - \frac{1}{K(a_2 + \gamma d)}\right)^n. \quad (8)$$

We note that for large $K$, we will have that

$$\frac{1}{K(a_2 + \gamma d)} < \frac{3}{4},$$

say.

**Proof** For $1 \leq k \leq n$, let

$$p_{n,k} := P(T(2) \leq k | T^{(1)}_{dK} > n), r_{n,k} := 1 - p_{n,k}.$$ 

For $i \geq 1$ and $j \geq dK$,

$$\left(\frac{i}{a_2 K + \gamma j + i}\right)^i \leq \left(\frac{i}{a_2 K + \gamma dK + i}\right)^i \leq \frac{1}{K(a_2 + \gamma d)}.$$
since the expression in the middle decreases with \( i \). Hence,

\[
P_{n,k} = p_{n,k-1} + \sum_{j=dK+1}^{\infty} \sum_{i=1}^{\infty} P(Z_{k-1}^{(1)} = j; Z_{k-1}^{(2)} = i \mid T_{dK}^{(1)} > n) \left( \frac{i}{a_2 K + \gamma j + i} \right)^i
\]

\[
\leq p_{n,k-1} + \sum_{j=dK+1}^{\infty} \sum_{i=1}^{\infty} P(Z_{k-1}^{(1)} = j; Z_{k-1}^{(2)} = i \mid T_{dK}^{(1)} > n) \frac{1}{K(a_2 + \gamma d)}
\]

\[
= p_{n,k-1} + \frac{1}{K(a_2 + \gamma d)} r_{n,k-1}.
\]

Thus, for any \( k \leq n \)

\[
r_{n,k} \geq \left( 1 - \frac{1}{K(a_2 + \gamma d)} \right) r_{n,k-1},
\]

which, evidently, implies

\[
r_{n,n} \geq \left( 1 - \frac{1}{K(a_2 + \gamma d)} \right)^n
\]

proving Equation (8).

A similar argument can be used to find an exponential bound for mutant survival probabilities, given that the resident does not surpass high levels [10].

Like the single trait process, considered in Section 2, the present process behaves as a deterministic dynamical system perturbed by small noise, which is, however, not to be forgotten.

Write the density process

\[
X_n^{(1)} = \frac{Z_n^{(1)}}{K}, \quad X_n^{(2)} = \frac{Z_n^{(2)}}{K}
\]

and note that the offspring distributions, as given in the beginning of this section, are again functions of the density. As before,

\[
X_{n+1}^{(1)} = \frac{1}{K} \sum_{j=1}^{\infty} \xi_j^{(1)}, \quad X_{n+1}^{(2)} = \frac{1}{K} \sum_{j=1}^{\infty} \xi_j^{(2)}.
\]

Denote the offspring mean by \( m(x) \) when the density is \( x \), \( m = (m_1, m_2) \),

\[
m_1(x) = E[\xi^{(1)}] = \frac{2a_1}{a_1 + x_1 + x_2}, \quad m_2(x) = E[\xi^{(2)}] = \frac{2a_2}{a_2 + \gamma x_1 + x_2}
\]

and write

\[
X_{n+1}^{(1)} = \frac{1}{K} \sum_{j=1}^{\infty} (\xi_j^{(1)} - m_1(X_n^{(1)}, X_n^{(2)})) + \underbrace{X_n^{(1)} m_1(X_n^{(1)}, X_n^{(2)})}_{f_1(X_n^{(1)}, X_n^{(2)})}, \quad (9)
\]

\[
X_{n+1}^{(2)} = \frac{1}{K} \sum_{j=1}^{\infty} (\xi_j^{(2)} - m_2(X_n^{(1)}, X_n^{(2)})) + \underbrace{X_n^{(2)} m_2(X_n^{(1)}, X_n^{(2)})}_{f_2(X_n^{(1)}, X_n^{(2)})}, \quad (10)
\]

The deterministic dynamics is given by

\[
f_1(x) = \frac{2x_1a_1}{a_1 + x_1 + \gamma x_2}, \quad f_2(x) = \frac{2x_2a_2}{a_2 + \gamma x_1 + x_2}, \quad (11)
\]
and

\[ x_{n+1}^{(1)} = f_1(x_n^{(1)}, x_n^{(2)}), \quad x_{n+1}^{(2)} = f_2(x_n^{(1)}, x_n^{(2)}). \]

It is also easy to see that the noise term when the system is at \( x \) is of order \( 1/\sqrt{K} \). Indeed,

\[ \eta^{(i)} = \frac{1}{\sqrt{K}} \sum_{j=1}^{Kw} (\xi_j^{(i)} - m_i), \quad i = 1, 2 \]

has mean zero and variance one. Hence, we can write Equations (9) and (10) as

\[ X_{n+1}^{(1)} = f_1(X_n^{(1)}, X_n^{(2)}) + \frac{1}{\sqrt{K}} \eta_n^{(1)}, \quad (12) \]

\[ X_{n+1}^{(2)} = f_2(X_n^{(1)}, X_n^{(2)}) + \frac{1}{\sqrt{K}} \eta_n^{(2)}. \quad (13) \]

The function \( f \) in Equation (11) that generates this system has two fixed points \( 0 \) and \( x^* = (x_1^*, x_2^*) \).

This is exactly the point at which both reproductions are critical. The determinant \( \Delta \) of partial derivatives shows that \( 0 \) is repelling, since \( \Delta = 4 \) there. At the point \( x^* \), the determinant of partial derivatives is

\[ \Delta = \begin{vmatrix} a_1 - 2a_1\gamma^2 + \gamma a_2 & \gamma(a_1 - \gamma a_2) \\ 2a_1(1 - \gamma^2) & 2a_1(1 - \gamma^2) \\ -\gamma(a_2 - \gamma a_1) & a_2 - 2a_2\gamma^2 + \gamma a_1 \\ 2a_2(1 - \gamma^2) & 2a_2(1 - \gamma^2) \end{vmatrix}, \]

which allows deriving conditions for stability \( \Delta < 1 \) of that fixed point in terms of the parameters \( \gamma, a_1, a_2 \).

Indeed, the deterministic system has four equilibrium points \( (0, 0), (a_1, 0), (0, a_2) \) and \( (x_1^*, x_2^*) \). As pointed out, the origin is always unstable. The fixed point corresponding to \( (x_1^*, x_2^*) \) lies in the positive quadrant only if \( \gamma \) is less than the maximum of \( a_2/a_1 \) and \( a_1/a_2 \). In that case, it is also attracting, and the other two points are unstable. If \( \gamma > a_1 \), the point \( (a_1, 0) \) is stable, so the second morph goes extinct. If \( \gamma > a_1/a_2, (0, a_2) \) is stable, and the former resident goes extinct.

4. A simple stochastic model with mutation and adaptive dynamics

Having investigated one lone species and the joint behaviour of a resident and a mutant, we return to the well-established special [6] model sketched in Section 1, determined by Gaussian functions.

It is controlled by five parameters: \( K, \mu, \sigma_1, \sigma_2 \) and \( \sigma_3 \), with the following interpretation.

- \( K \) is the maximal carrying capacity of a monomorphic population (large).
- \( \mu \) is the mutation rate per gene per generation (and very small).
- \( \sigma_1^2 \) describes how the carrying capacity varies.
- \( \sigma_2^2 \) is often referred to as the competition width.
- \( \sigma_3^2 \) is the variance of mutational steps (again small).
Important compound parameters are

$$\beta = \frac{\sigma^2_1}{\sigma^2_2} \text{ and } \delta = \frac{\sigma^2_1 - \sigma^2_2}{\sigma^2_1 + \sigma^2_2}.$$ 

Note that $-1 < \delta < 1$,

$$1 - \delta = \frac{2\sigma^2_2}{\sigma^2_1 + \sigma^2_2}, \hspace{0.5cm} 1 + \delta = \frac{2\sigma^2_1}{\sigma^2_1 + \sigma^2_2},$$

and

$$\frac{1 - \delta}{1 + \delta} = \frac{\sigma^2_2}{\sigma^2_1} = \beta^{-1}.$$ 

A morph with trait $t \in \mathbb{R}$ has the carrying capacity

$$K(t) = Ke^{-t^2/2\sigma_1^2},$$ (14)

explaining the role of $\sigma_1^2$, in what sense it describes how carrying capacity varies for different traits, with the maximum $K(0) = K$ for the trait value $t = 0$.

In a polymorphic population, the competition between two morphs $t_1$ and $t_2$ is quantified by the coefficient

$$\gamma(t_1, t_2) = e^{-(t_1-t_2)^2/2\sigma_1^2},$$ (15)

depending only on the difference $|t_1 - t_2|$ and reaching its maximum value $\gamma(t, t) = 1$ for individuals with the same trait value.

Mutation in a monomorphic population occurs at a rate of the order $\mu$ mutations per individual and time unit. From the preceding analysis, the mutation rate should be taken to satisfy $e^{-cK} \ll \mu \ll 1/K$, the constant $c$ defined in Equation (3), so as to avoid repeated mutations during the time while one mutant establishes itself, while guaranteeing mutations during the residence. In adaptive dynamics, this would be formulated as conditions on mutations not occurring in ecological time spans, but (almost) certainly in an evolutionary time scale. Given a mutation, the new trait value $t_2$ then has the normal conditional density function

$$\phi(t_2 | t_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(t_1-t_2)^2/2\sigma^2},$$ (16)

implying symmetry in the step distribution and that larger changes in trait value are less probable.

The process starts from a finite number of individuals of a finite number of different traits. Traits in the various generations are denoted by vectors $(t_1, t_2, \ldots, t_p)$, where $p$ can vary from generation to generation. Denote the number of individuals with trait $t$ in the $n$th generation by $Z_n(t)$. Given the vector $(Z_n(t_1), Z_n(t_2), \ldots, Z_n(t_p))$, individuals of the $n$th generation reproduce independently, splitting into two with the probability

$$\frac{K(t)}{K(t) + \sum_j \gamma(t, t_j)Z_n(t_j)} \text{ for } t = t_1, t_2, \ldots, t_p,$$

otherwise begetting no child. Each child is a mutant with probability $\mu$, the mutant’s trait being normally distributed around its mother’s, and with the variance $\sigma^2$. 

This is an informal (but precise!) description of a discrete-time stochastic process, whose states are counting measures on $\mathbb{R}$. If it starts from one individual at say $t_1$, it will develop according to Section 2 (with $K(t_1)$ replacing $K$), until the first mutation occurs. Then we find ourselves in the circumstances of Section 3 with $\gamma$ as defined by Equation (15), $a_1 = \exp(-t_1^2/(2\sigma_1^2))$, and $a_2 = \exp(-t_2^2/(2\sigma_1^2))$, $K = K(0)$, according to Equation (14).
After some algebra from the inequality $a_2 > \gamma a_1$ (Equation (4)), we arrive at a simple condition for the mutants possibly being able to establish themselves: If $t_2 < t_1$, then also $t_2 < \delta t_1$. If $t_1 > 0$, the second condition is empty, whereas if $t_1 < 0$, $t_2$ must be at a considerable distance left of $t_1$, which is not feasible due to the smallness of $\sigma$. We conclude that residents with traits on the right half-axis can only establish mutants to the left of them, and vice versa. Indeed, the situation is summarized by the condition

$$(t_2 - t_1)(t_2\delta - t_1) > 0$$

being necessary for establishment, whatever be the position of $t_1$, and equivalent to $t_2/t_1 \in (-\infty, 1) \cup (1/\delta, \infty)$.

This means that in the monomorphic phase, the trait value will move towards zero until it reaches a $\sigma$-neighbourhood of zero. Then the trait value may jump in both directions and moreover the first evolutionary branching becomes plausible. Indeed, the reverse invasion condition is

$$(t_1 - t_2)(t_1\delta - t_2) > 0$$

equivalent to $t_2/t_1 \in (-\infty, \delta) \cup (1, \infty)$. When both conditions are fulfilled, the pair of trait values $(t_1, t_2)$ may coexist.

In a slightly more precise wording, we start the process from a monomorphic population, say with trait $T_0 = t > 0$. The population will either die out or (quasi-)establish itself at its carrying capacity. In the latter case, sooner or later at a random time $\nu_1$, a successful mutant will replace it, and the trait will have moved to $T_{\nu_1}$, where it will stay until the next successful mutation $\nu_2$, until finally it would have come so close to zero, so as to render branching possible, the trait being of the order of $\sigma$. Then $\{T_n\}$ will take the value of two traits. Thus, we obtain a stochastic process with values in $\bigcup_n \mathbb{R}^n$, about which many interesting questions can be posed.

1. What is the time to the first successful mutation? In other words, what is the distribution of $\nu_1$?
2. When will the first branching occur, or rather, what is the distribution of the waiting time until a monomorphic population turns dimorphic, the distribution to the next change in number of morphs, etc.
3. How can the stochastic movement $\{T_n\}$ in trait space be described?
4. And how the degree of polymorphism (the development of the number of coexisting species)?
5. What limiting behaviours can be discerned, when $K \to \infty$ and $\mu \to 0$?

For birth-and-death processes, many of these matters have, as pointed out, been discussed by Méléard, Champagnat et al. (see in particular [3]). Our more mathematical findings on the present simple model will appear in [10]. As one of the referees pointed out, the basic method used here, domination by not-size-dependent branching processes during time periods when populations sizes satisfy certain inequalities, invites usage also in more realistic, age-dependent or general branching processes.

Acknowledgements

The paper has profited from a knowledgeable and dedicated refereeing. We thank the three referees. F.C. Klebaner was supported by the Australian Research Council Grant DP0881011. S. Sagitov was supported by the Bank of Sweden Tercentenary Foundation. V.A. Vatutin was supported by the RAS programme ‘Mathematical Control Theory’ and INTAS 03-51-5018. P. Haccou was supported by the Netherlands Organisation for Scientific Research (NWO). P. Jagers was supported by the Swedish Research Council.

References

[1] F. Bugge Christiansen and V. Loeschke, *Evolution and intraspecific competition. I. One-locus theory for small additive gene effects*, Theor. Popul. Biol. 18 (1980), pp. 297–313.
[2] N. Champagnat and A. Lambert, *Adaptive dynamics in logistic branching population*, Banach Center Publ. Polish Acad. Sci. 80 (2008), pp. 235–244.

[3] N. Champagnat and S. Méléard, *Polymorphic evolution sequence and evolutionary branching*, Th. Prob. Rel. Fields (2010), to appear.

[4] N. Champagnat, R. Ferrière, and S. Méléard, *From individual stochastic processes to macroscopic models in adaptive evolution*, Stoch. Models 24(Suppl. 1) (2008), pp. 2–44.

[5] U. Dieckmann and R. Law, *The dynamical theory of coevolution: A derivation from stochastic ecological processes*, J. Math. Biol. 34 (1996), pp. 579–612.

[6] U. Dieckmann and M. Doebeli, *On the origin of species by sympatric speciation*, Nature 400 (6742) (1999), pp. 354–357.

[7] M. Durinx, J.A.J. Metz, and G. Meszéna, *Adaptive dynamics for physiologically structured models*, J. Math. Biol. 56 (2008), pp. 673–742.

[8] P. Haccou, P. Jagers, and V.A. Vatutin, *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge University Press, Cambridge, 2005.

[9] P. Jagers, *Extinction, persistence, and evolution*, in *The Mathematics of Darwin’s Legacy*, F.A. Chalub, da Costa, and J.F. Rodrigues, eds., Birkhäuser, Basel, 2011.

[10] P. Jagers, F.C. Klebaner, V.A. Vatutin, and S. Sagitov, *The mathematics of a stochastic adaptive dynamics model of evolution*, to appear.

[11] S. Janson, *Large deviation inequalities for sums of indicator variables*, Tech. Rep., Department of Mathematics, Uppsala University, 1994, p. 34.

[12] Y. Kifer, *Random Perturbations of Dynamical Systems*, Birkhäuser, Boston, 1988.

[13] F.C. Klebaner, *On population-size-dependent branching processes*, Adv. Appl. Probab. 16 (1984), pp. 30–55.

[14] F.C. Klebaner, *Population-dependent branching processes with a threshold*, Stoch. Process Appl. 46 (1993), pp. 115–127.

[15] F.C. Klebaner and J. Lazar, *On the quasi-stationary distribution in randomly perturbed dynamical systems with a single attracting point*, in Proceedings of 2nd Australia–Japan Workshop on Stochastic Models in Engineering, Technology and Management, R.J. Wilson, D.N. Pra Murthy, and S. Osaki, eds., The University of Queensland, Brisbane, 1996, pp. 348–354.

[16] F.C. Klebaner and O. Nerman, *Autoregressive approximation in population-dependent branching processes with a threshold*, Stoch. Process. Appl. 51 (1994), pp. 1–7.

[17] F.C. Klebaner and O. Zeitouni, *The exit problem for a class of density dependent branching systems*, Ann. Appl. Probab. 4 (1994), pp. 1188–1205.

[18] F.C. Klebaner, J. Lazar, and O. Zeitouni, *On the quasi-stationary distribution of some randomly perturbed transformations of the interval*, Ann. Appl. Probab. 8 (1998), pp. 300–315.

[19] P.H. Leslie, *A stochastic model for studying the properties of certain biological systems by numerical methods*, Biometrika 55 (1958), pp. 16–31.

[20] J.A.J. Metz, R.M. Nisbet, and S.A.H. Geritz, *How should we define ‘fitness’ for general ecological scenarios?*, TREE 7 (1992), pp. 198–202.

[21] J.A.J. Metz, S.A.H. Geritz, G. Meszna, F.J.A. Jacobs, and J.S. van Heerwaarden, *Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction*, in *Stochastic and Spatial Structures of Dynamical Systems*, S.J. van Strien and S.M. Verduyn Lunel, eds., North-Holland, Amsterdam, 1996, pp. 183–231.

[22] O. Nerman, *On the maximal generation size of a non-critical Galton–Watson process*, Scand. J. Statist. 4 (1977), pp. 131–135.

[23] H.A. Orr, *The distribution of fitness effects among beneficial mutations*, Genetics 163 (2003), pp. 1519–1526.

[24] K. Parvinen, *Evolutionary suicide*, Acta Biotheor. 53 (2005), pp. 241–264.

[25] K. Ramanan and O. Zeitouni, *The quasi-stationary distribution for small random perturbations of certain one-dimensional maps*, Stoch Process. Appl. 84 (1999), pp. 25–51.

[26] J. Roughgarden, *Resource partitioning among competing species – A coevolutionary approach*, Theor. Popul. Biol. 9 (1976), pp. 388–424.

[27] J. Roughgarden, *Theory of Population Genetics and Evolutionary Ecology: An Introduction*, MacMillan, New York, 1979.