A statistical model of an evolving population with sexual reproduction

1991 J. Phys. A: Math. Gen. 24 L705

(http://iopscience.iop.org/0305-4470/24/13/001)

J. Phys. A: Math. Gen. 24 (1991) L705-L709. Printed in the UK

LETTER TO THE EDITOR

A statistical model of an evolving population with sexual reproduction

M Serva† and L Peliti‡

† Dipartimento di Matematica, Università dell'Aquila, I-67010 Coppito, Italy
‡ Dipartimento di Scienze Fisiche and Unità INFM, Università 'Federico II', Mostra d'Oltremare, Pad 19, I-80125 Napoli, Italy||

Received 19 March 1991, in final form 9 May 1991

Abstract. We introduce a statistical model of a population with a sexual reproduction mechanism, evolving in a flat fitness landscape. We show that fluctuations in the genetic overlap between different individuals in the population vanish in the infinite-population limit, contrary to the case of asexual reproduction, which exhibited spin-glass-like behaviour.

A statistical model of a population evolving in the absence of natural selection (or, if one uses the useful metaphor of Sewall Wright, in a flat fitness landscape) has been recently introduced and solved (Derrida and Peliti 1990, hereafter denoted by DP) by exploiting the analogy with a dynamical model with stochastic dynamics: the annealed random map model (Derrida and Bessis 1988). One of the most interesting features of the solution was its analogy with spin glasses (for a review, see Mézard *et al* 1987). The distribution of the population in genetic space could be described by a quantity analogous to the overlap distribution in spin glasses $P(q)$. It turned out that this quantity exhibited fluctuations even in a suitably defined infinite-population limit. One could therefore speak of two averages: the *population average* is the analogue of the thermal average in disordered systems, while the *process average*, i.e. the average over all possible realizations of the reproduction process (conveniently represented in the simulations by a time average), is the analogue of the average over disorder.

We consider in this letter a model similar to the one just discussed, except that the reproduction mechanism is analogous to sexual reproduction. We have exploited the privilege of defining the model in order to keep its average properties as similar as possible to that one, in such a way as to facilitate comparison. Our main result is that *the genetic distance between different individuals does not fluctuate in the infinite-population limit.* In such a way most of the analogy with disordered systems appears to be irrelevant in sexually reproducing populations evolving in a flat fitness landscape. Since natural selection (or a non-trivial fitness landscape) has in general the tendency to reduce fluctuations, it is expected that fluctuations of average properties of a sexually reproducing population should be rather smaller than for an asexually reproducing one (for an accessible reference on evolutionary genetics, see Maynard Smith (1989)).

§ Associato INFN, Sezione dell'Aquila.
|| Associato INFN, Sezione di Napoli.

0305-4470/91/130705+05$03.50

We consider a population formed by a fixed number, M, of individuals, the genetic structure of which is identified by the state of N binary units, $\sigma_i^\alpha = \pm 1$, $i = 1, \ldots, N$, $\alpha = 1, \ldots, M$. The population evolves by the combined effects of reproduction and mutation, but in the absence of natural selection: i.e. we assume that the probability of reproductive success of an individual is independent of its genetic structure.

We define two models, distinguished by the reproduction mechanism. The model with asexual reproduction, introduced and solved by DP, will be denoted as model A; the model with sexual reproduction, whose discussion is the subject of this letter, will be denoted as model B.

In model A, given the population at some generation t, the following generation $t+1$ is obtained as follows: for each individual α one chooses at random, independently for each individual and with uniform probability, the parent $\beta = G_t(\alpha) \in \{1, \ldots, M\}$. The generic structure of the new individual α is then identical to that of its parent, up to mutations, which occur with probability $\mu\, dt$ in the short time interval dt. One has therefore, for each unit $i = 1, \ldots, N$ in the genome:

$$\sigma_i^\alpha(t+1) = \begin{cases} \sigma_i^\beta(t) & \text{with probability } \frac{1}{2}(1+e^{-2\mu}) \\ -\sigma_i^\beta(t) & \text{with probability } \frac{1}{2}(1-e^{-2\mu}). \end{cases} \tag{1}$$

In model B, at each generation and for each individual α, one chooses independently and with uniform probability *two* parents, $\alpha' = G_t^{(1)}(\alpha)$ and $\alpha'' = G_t^{(2)}(\alpha)$, where $\alpha' \neq \alpha''$. For each unit i and each individual α, one then chooses with equal probability one of the two parents, and the state of the unit is equal to that of the corresponding parent, up to mutations which occur with the same probability as before. One thus has the same equation (1) as before, but now, for each individual α, each unit i and each generation t, one chooses independently and with equal probability $\beta = \alpha'$ or $\beta = \alpha''$. This model obviously neglects linkage, whereas the presence of mutations allows it to move away from simple Hardy-Weinberg equilibrium.

Both models can be conveniently represented by stochastic equations. The effect of mutations can be represented by introducing, for each unit i, individual α and generation t, an independent random variable $\varepsilon_i^\alpha(t)$, which takes the value $+1$ with probability $\frac{1}{2}(1+e^{-2\mu})$ and the value -1 with probability $\frac{1}{2}(1-e^{-2\mu})$. The state of unit (i, α) at generation $(t+1)$ is then given, in model A, by the expression

$$\sigma_i^\alpha(t+1) = \varepsilon_i^\alpha(t)\sigma_i^{\alpha'}(t) \tag{2}$$

where $\alpha' = G_t(\alpha)$ is the parent of individual α at generation $(t+1)$. By the same token, one introduces in model B the independent random variables ξ_i^α, which take the values 0 or 1 with probability $\frac{1}{2}$. One then has, in model B,

$$\sigma_i^\alpha(t+1) = \varepsilon_i^\alpha(t)[\xi_i^\alpha(t)\sigma_i^{\alpha'}(t) + (1+\xi_i^\alpha(t))\sigma_i^{\alpha''}(t)] \tag{3}$$

where $\alpha' = G_t^{(1)}(\alpha)$ and $\alpha'' = G_t^{(2)}(\alpha)$ are the two parents of individual α at generation $(t+1)$.

As pointed out by DP, it is in general necessary to consider two different kinds of average: the *population average*, denoted by angular brackets $\langle\ \rangle$, and the *process average*, denoted by bar $\overline{\ }$. For example, the genetic similarity between two individuals α and β may be represented by their overlap $q^{\alpha\beta}$, defined by

$$q^{\alpha\beta} = \frac{1}{N}\sum_i \sigma_i^\alpha \sigma_i^\beta. \tag{4}$$

Therefore, the genetic similarity of the population as a whole may be represented by the population average $Q=\langle q\rangle$ of the overlap:

$$Q=\langle q\rangle=\left[\binom{M}{2}\right]^{-1}\sum_{(\alpha,\beta)}q^{\alpha\beta} \tag{5}$$

where the sum runs over all different pairs of individuals in the population. (We remark that this definition differs slightly from that introduced in DP). However, Q fluctuates in general from generation to generation. One thus introduces its process average $\bar{Q}$, whose value can be calculated as follows. In model A one has, from (2) and for $\alpha\neq\beta$,

$$\overline{\sigma_i^{\alpha}(t+1)\sigma_i^{\beta}(t+1)}=\overline{\varepsilon_i^{\alpha}(t)\varepsilon_i^{\beta}(t)\sigma_i^{\alpha'}(t)\sigma_i^{\beta'}(t)}=\mathrm{e}^{-4\mu}\,\overline{\sigma_i^{\alpha'}(t)\sigma_i^{\beta'}(t)}. \tag{6}$$

Now $\overline{\sigma_i^{\alpha'}(t)\sigma_i^{\beta'}(t)}$ is equal to 1 if $\alpha'=\beta'$ (which happens with probability $1/M$) and to $\overline{Q(t)}$ otherwise, because of the equivalence of all units i and pairs of individuals (α,β). We obtain therefore

$$\overline{Q(t+1)}=\mathrm{e}^{-4\mu}\left[\frac{1}{M}+\left(1-\frac{1}{M}\right)\overline{Q(t)}\right] \tag{7}$$

and in the steady state where $\overline{Q(t+1)}=\overline{Q(t)}=\bar{Q}$,

$$\bar{Q}=\frac{1}{M(\mathrm{e}^{4\mu}-1)+1}. \tag{8}$$

Let us now turn to model B and equation (3). We have

$$\begin{aligned}\overline{\sigma_i^{\alpha}(t+1)\sigma_i^{\beta}(t+1)}&=\mathrm{e}^{-4\mu}\frac{1}{4}\overline{[\sigma_i^{\alpha'}(t)+\sigma_i^{\alpha''}(t)][\sigma_i^{\beta'}(t)+\sigma_i^{\beta''}(t)]}\\&=\mathrm{e}^{-4\mu}\frac{1}{4}\left[\frac{4}{M}+4\left(1-\frac{1}{M}\right)\overline{Q(t)}\right].\end{aligned} \tag{9}$$

We obtain therefore the same equation (7) and the same steady-state result (8) as in model A. These results become simpler in the limit $m\to\infty$, $\mu M=$ constant, considered by DP. One has in fact $\bar{Q}=\lambda/(1+\lambda)$, where

$$\lambda=\frac{1}{4\mu M}. \tag{10}$$

Although the average variability of the two models is the same, the corresponding fluctuations are different. To proceed further, it is useful to take, along with the infinite-population limit defined above, the infinite-genome limit (Kimura 1983), $N\to\infty$. One thus obtains:

$$A=\bar{Q}=\overline{\langle q\rangle}=\overline{\sigma_i^{\alpha}\sigma_i^{\beta}} \tag{11}$$

$$B=\overline{\langle q^2\rangle}=\overline{\sigma_i^{\alpha}\sigma_i^{\beta}\sigma_j^{\alpha}\sigma_j^{\beta}} \tag{12}$$

$$C=\overline{Q^2}=\overline{\langle q\rangle^2}=\overline{\sigma_i^{\alpha}\sigma_i^{\beta}\sigma_j^{\gamma}\sigma_j^{\delta}}. \tag{13}$$

These relations hold up to terms of order $1/M$ and $1/N$, and it is understood that different Latin indices refer to different genome units, and likewise different Greek indices refer to different individuals. In addition to the quantities defined above, it is useful to introduce the quantity D, defined by

$$D=\overline{\sigma_i^{\alpha}\sigma_j^{\alpha}\sigma_i^{\beta}\sigma_j^{\gamma}}. \tag{14}$$

We can now compute the fluctations of the overlap. In model A, by means of (2), we obtain:

$$C(t+1)=\overline{\sigma_i^{\alpha}(t+1)\sigma_i^{\beta}(t+1)\sigma_j^{\gamma}(t+1)\sigma_j^{\delta}(t+1)}$$
$$=e^{-8\mu}\,\overline{\sigma_i^{\alpha'}(t)\sigma_i^{\beta'}(t)\sigma_j^{\gamma'}(t)\sigma_j^{\delta'}(t)}. \tag{15}$$

The four individuals α', β', γ', δ' are all different with probability equal to $(M-1)\times(M-2)(M-3)/M^3\simeq 1-(6/M)$. With probability $2/M$ one has $\alpha'=\beta'$ or $\gamma'=\delta'$ and with probability $4/M$ one has $\alpha'=\gamma'$ or $\alpha'=\delta'$ or $\beta'=\gamma'$ or $\beta'=\delta'$. Other possible combinations yield contributions of order $1/M^2$, which are negligible. This implies the following steady-state equation:

$$C=e^{-8\mu}\left[\left(1-\frac{6}{M}\right)C+\frac{2}{M}A+\frac{4}{M}D\right]. \tag{16}$$

One also analogously obtains:

$$B(t+1)=e^{-8\mu}\left[\frac{1}{M}+\left(1-\frac{1}{M}\right)B(t)\right] \tag{17}$$

$$D(t+1)=e^{-8\mu}\left[\frac{2}{M}A(t)+\frac{1}{M}B(t)+\left(1-\frac{3}{M}\right)D(t)\right]. \tag{18}$$

These expressions imply, beyond (8),

$$B=\frac{\lambda}{\lambda+2}$$
$$C=\frac{\lambda^2(9\lambda^2+18\lambda+4)}{(\lambda+1)(\lambda+2)(3\lambda+1)(3\lambda+2)} \tag{19}$$
$$D=\frac{5\lambda^2+3\lambda^3}{(\lambda+1)(\lambda+2)(3\lambda+2)}.$$

These results confirm the expression of the fluctuations of Q derived by DP.

Turning now to model B, we have, neglecting terms of order $1/M^2$:

$$B(t+1)=\overline{\sigma_i^{\alpha}(t+1)\sigma_i^{\beta}(t+1)\sigma_j^{\alpha}(t+1)\sigma_j^{\beta}(t+1)}$$
$$=\frac{e^{-8\mu}}{16}\overline{(\sigma_i^{\alpha'}+\sigma_i^{\alpha''})(\sigma_i^{\beta'}+\sigma_i^{\beta''})(\sigma_j^{\alpha'}+\sigma_j^{\alpha''})(\sigma_j^{\beta'}+\sigma_j^{\beta''})}$$
$$=\frac{e^{-8\mu}}{16}\left[4\left(1-\frac{4}{M}\right)(B+C+2D)+\frac{4}{M}(1+6A+3B+6D)\right]. \tag{20}$$

The quantities on the last line are evaluated at generation t. Here, in contrast with (17), the leading terms are of order one, instead of order $1/M$. It is sufficient therefore to consider just the case in which all four parents are different. Neglecting terms of order $1/M$, we obtain in the DP limit the following steady-state equation:

$$3B=C+2D. \tag{21}$$

In an analogous way the following equations for C and D can be derived:

$$C=e^{-8\mu}\left[\left(1-\frac{6}{M}\right)C+\frac{2}{M}A+\frac{4}{M}D\right] \tag{22}$$

$$D=\frac{e^{-8\mu}}{16}\left(8D+8C+\frac{4}{M}A\right). \tag{23}$$

The last equation implies, in the DP limit, $D = C$, which, together with equation (21), implies $B = C$. Replacing these values in equation (22) we obtain $B = C = D = A^2 = 1/(1+1/\lambda)^2$, leading to

$$\overline{\langle q\rangle^2} = \overline{\langle q^2\rangle} = \overline{\langle q\rangle}^2. \tag{24}$$

This surprising result implies that, in the infinite-population limit, the situation is very different from the one holding in the asexual reproduction case. There, the overlap was distributed in the population according to the wider or closer relatedness of two individuals—yielding a non-trivial overlap distribution $P(q)$—and moreover this distribution itself did fluctuate from generation to generation. Here, the average overlap Q does not fluctuate, and, more strikingly, *the relative overlap of any two individuals is equal to the average with probability one.* This is due to the fact that the ancestors (t generations ago) of each individual sample, whenever t exceeds a few times $\log_2 M$, the whole population over and over again, effectively reducing the fluctuations.

On the other hand, the average properties of the two models coincide. Beyond the average overlap Q, it is indeed also possible to evaluate the effective mutation rate μ^*, defined by the behaviour of the correlation function

$$\chi(t) = \frac{1}{N}\sum_i \overline{\langle\sigma_i(t)\rangle\langle\sigma_i(0)\rangle} \propto e^{-2\mu^*|t|}. \tag{25}$$

By taking the process average of (3) it is easy to see that

$$\overline{\sigma_i^{\alpha}(t+1)} = \frac{e^{-2\mu}}{2}\,\overline{(\sigma_i^{\alpha'}(t)+\sigma_i^{\alpha''}(t))} = e^{-2\mu}\overline{\sigma_i^{\alpha'}(t)} \tag{26}$$

yielding

$$\mu^* = \mu \tag{27}$$

as in DP, confirming a famous result of Kimura (1983).

We have thus shown that, although the average properties of model B (with sexual reproduction mechanism) are equal to those of model A, its fluctuations are radically different, and are negligible in the infinite-population limit. This vindicates the relevance of fluctuations in small populations as the source of evolutionary innovation.

Note added. B Derrida and P Higgs (private communication) have recently considered a model similar to ours, reaching the same results. However, if it is assumed that the coupling between individuals α and β is only fecund if the overlap $q^{\alpha\beta}$ is larger than a threshold q_0, they find that the probability distribution $p(q)$ of the overlap becomes non-trivial whenever q_0 exceeds $\lambda/(1+\lambda)$. Indeed, in this situation the population breaks up in mutually unfecund sub-populations (species), whose size fluctuates rapidly as time goes on. It appears that in this language, speciation is manifested by the appearance of a non-trivial Parisi order parameter, i.e. by a spin-glass phase transition.

References

Derrida B and Bessis D 1988 *J. Phys. A: Math. Gen.* **21** L509–15
Derrida B and Peliti L 1990 *Bull. Math. Biol.* in press
Kimura M 1983 *The Neutral Theory of Molecular Evolution* (Cambridge: Cambridge University Press)
Maynard Smith J 1989 *Evolutionary Genetics* (Oxford: Oxford University Press)
Mézard M, Parisi G and Virasoro M A 1987 *Spin-Glass Theory and Beyond* (Singapore: World Scientific)