Electronic Supplementary Material for Error catastrophe in populations under similarity-essential recombination

Marcus A.M. de Aguiar[1], David M. Schneider[1], Eduardo do Carmo[2], Paulo R. A. Campos[3], Ayana B. Martins[1,4]

[1]*Instituto de Física 'Gleb Wataghin', Universidade Estadual de Campinas, 13083-970, Campinas, SP, Brazil*

[2]*Universidade Federal da Integração Latino Americana 85867-970, Foz do Iguaçu, PR, Brazil*

[3]*Departamento de Física, Universidade Federal de Pernambuco, 50670-901, Recife, PE, Brazil*

[4]*Instituto de Biociências, Universidade de São Paulo, 05508-090, São Paulo, SP, Brazil*

I. NORMALIZATION

We first verify that $S_{j,k} \equiv \sum_i c_\mu(j,k;i) = 1$:

$$\begin{aligned} S_{j,k} &\equiv \sum_i (1-\mu)^{B-\alpha-\gamma}\mu^\alpha \left(\frac{1}{2}\right)^\gamma \\ &= (1-\mu)^{B-\gamma}\left(\frac{1}{2}\right)^\gamma \sum_i \left(\frac{\mu}{1-\mu}\right)^\alpha \end{aligned} \tag{1}$$

where $\gamma = d(j,k)$. The number of loci bearing the same allele in j and k is $B-\gamma$ and α is the subset of these loci where the alleles differ from that in i. We can therefore replace the sum over i by a sum over α running from 0 to $(B-\gamma)$. For each α there are $\binom{B-\gamma}{\alpha}$ possible permutations for the part of i where the alleles of j and k coincide and 2^γ combinations for the rest. We obtain

$$S_{j,k} = (1-\mu)^{B-\gamma}\left(\frac{1}{2}\right)^\gamma \sum_{\alpha=0}^{B-\gamma} \begin{pmatrix} B-\gamma \\ \alpha \end{pmatrix} 2^\gamma \left(\frac{\mu}{1-\mu}\right)^\alpha = 1. \tag{2}$$

Summing the dynamical equation

$$p_i^{t+1} = \sum_{j,k,\gamma\le G} p_j^t\, p_k^t\, c_\mu(j,k;i). \tag{3}$$

over i on both sides we find

$$\begin{aligned} \sum_i p_i^{t+1} &= \sum_{i,j,k} p_j^t\, p_k^t\, (1-\mu)^{B-\alpha-\gamma}\mu^\alpha\left(\frac{1}{2}\right)^\gamma = \sum_{j,k} p_j^t\, p_k^t S_{j,k} \\ &= \sum_j p_j^t \sum_k p_k^t = 1. \end{aligned} \tag{4}$$

A very similar demonstration holds for the case of assortative mating.

II. UNCONSTRAINED RECOMBINATION

When $G = B$ recombination is possible between every pair of individuals. In this case, if $\mu = 0$ the allele frequencies remain constant from the first generation and the haplotype frequencies asymptotically reach the linkage equilibrium. For $\mu \neq 0$ the haplotypes converge to the uniform distribution, and thereby all frequencies are equal to $p_i = 2^{-B}$.

The frequency ρ_{n,i_n} of the allele i_n in locus n is obtained by summing over all other alleles the haplotype frequencies of all types carrying i_n

$$\rho_{n,i_n} = \sum_{i_1 \ldots i_{n-1}, i_{n+1}, \ldots i_B} p_{i_1 \ldots i_{n-1} i_n i_{n+1} \ldots i_B}, \tag{5}$$

where we have written explicitly $p_i = p_{i_1 i_2 \ldots i_B}$. Also $i_n = 0$ or $i_n = 1$ and $\rho_{n,0} + \rho_{n,1} = 1$. Since alleles in each locus are segregated independently and there are no correlations between them, we can perform these sums in both sides of equation (8) in section II.A of the paper and obtain

$$\rho_{n,i_n}^{t+1} = \sum_{j_n=0}^{1} \sum_{k_n=0}^{1} \rho_{n,j_n}^t \; \rho_{n,k_n}^t \; (1-\mu)^{1-\alpha_1-\gamma_1} \mu^{\alpha_1} \left(\frac{1}{2}\right)^{\gamma_1} \tag{6}$$

where γ_1 and α_1 are given by equations (6) and (7) of the paper with $B = 1$. Explicitly,

$$\begin{aligned} \rho_{n,i_n}^{t+1} &= (\rho_{n,i_n}^t)^2 (1-\mu) + \rho_{n,i_n}^t (1 - \rho_{n,i_n}^t) + \mu (1 - \rho_{n,i_n}^t)^2 \\ &= \rho_{n,i_n}^t (1 - 2\mu) + \mu. \end{aligned} \tag{7}$$

whose solution is

$$\rho_{n,i_n}^t = (\rho_{n,i_n}^0 - 1/2)(1-2\mu)^t + 1/2. \tag{8}$$

This implies that the frequencies do not change if $\mu = 0$, $\rho_{n,i_n}^t = \rho_{n,i_n}^0$, which is the Hardy-Weinberg equilibrium for haploid individuals. In this case the only equilibrium solution for the haplotypes is the linkage equilibrium $p_{i_1,i_2,\ldots,i_B} = \rho_{1,i_1} \rho_{2,i_2} \ldots \rho_{B,i_B}$.

For $\mu \neq 0$ the allele frequencies converge to $\rho_{n,0} = \rho_{n,1} = 1/2$, a uniform distribution, where each allele is equally likely to be found in an individual. The haplotype frequencies, therefore, are given by $p_i = 2^{-B}$ and the error threshold is $\mu_c = 0$. An alternative proof that the uniform distribution is always a solution for unconstrained uniform crossover, along with the study of its stability, is presented here in sections III and IV.

III. UNIFORM DISTRIBUTION FOR UNRESTRICTED MATING

Here we prove that the uniform distribution $p_i = 2^{-B}$ is always a solution of the dynamical equations in the case of unrestricted mating, $G = B$.

Substituting $p_j^t = p_k^t = 2^{-B}$ in equation (3) we obtain

$$p_i^{t+1} = 2^{-2B} \sum_{j,k} (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma. \tag{9}$$

In order to performe the double sum we reorganize it in order of increasing $\gamma = d(j,k)$ for a fixed i:

- $\gamma = 0$. These terms have $j = k$ and $\alpha = d(i,j)$. They contribute to the sum with

$$\sum_j (1-\mu)^{B-\alpha}\mu^\alpha.$$

The sum is now organized in order of increasing α. For $\alpha = 0$ there is only one term with $j = i$. For $\alpha = 1$, j differs from i in only one gene, that can be $i_1, i_2, \dots, i_B$ so that there are B such terms. It is not difficult to see that for a given α the number of terms is given by the binomial

$$\begin{pmatrix} B \\ \alpha \end{pmatrix} = \frac{B!}{(B-\alpha)!\alpha!}.$$

We obtain

$$\sum_{\alpha=0}^{B} \begin{pmatrix} B \\ \alpha \end{pmatrix} (1-\mu)^{B-\alpha}\mu^\alpha = 1.$$

- $\gamma = 1$. Now $i'_k = i''_k$ for all k except one, for example for $k = B$. The two possibilities are $i'_B = 1, i''_B = 0$ or $i'_B = 0, i''_B = 1$. In either case we obtain $\alpha = \sum_{k=1}^{B-1} |i_k - i'_k|$, that can take up the values from 0 to $B-1$. The total number of terms of this type is $2B$ and their contribution to the sum is

$$\begin{aligned} &2B\sum_j (1-\mu)^{B-\alpha-1}\mu^\alpha \frac{1}{2} \\ &= B\sum_{\alpha=0}^{B-1} \begin{pmatrix} B-1 \\ \alpha \end{pmatrix} (1-\mu)^{B-\alpha-1}\mu^\alpha = B \end{aligned}$$

Continuing this way we see that the terms with a fixed γ contribute to the sum with the binomial $\begin{pmatrix} B \\ \gamma \end{pmatrix}$, so that

$$\sum_{j,k} (1-\mu)^{B-\alpha-\gamma}\mu^\alpha \left(\frac{1}{2}\right)^\gamma = \sum_{\gamma=0}^{B} \begin{pmatrix} B \\ \gamma \end{pmatrix} = 2^B. \tag{10}$$

Using this result in eq. (9) we obtain $p_i^{t+1} = 2^{-B}$, proving our assertion.

IV. STABILITY OF THE UNIFORM DISTRIBUTION

To study the stability of the uniform distribution in the case of unrestricted mating we linearize Eq.(3) around 2^{-B}. Setting

$$p_i = 2^{-B} + \delta p_i \tag{11}$$

and keeping only first order terms we obtain

$$\delta p_i^{t+1} = 2^{1-B} \sum_{j,k} \delta p_j^t \, (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma \equiv \sum_j J_{i,j} \, \delta p_j^t \, . \tag{12}$$

Notice that the Jacobian J is a large matrix with $2^B \times 2^B$ elements. The calculation of $J_{i,j}$ is performed in detail below and the result is

$$J_{i,j} = 2^{1-2B} \, (3-2\mu)^{B-d(i,j)} (1+2\mu)^{d(i,j)}. \tag{13}$$

The matrix elements depend only on $d(i,j)$ and only $B+1$ of them are independent. There are $2^B - 1$ eigenvalues $|\lambda_i| < 1$ whose eigenvectors satisfy $\sum_j v_{ij} = 0$ and one eigenvalue $|\lambda_1| > 1$ corresponding to the eigenvector $v = (1, 1, \ldots, 1)$. However, perturbations in the latter direction are not allowed, since the expansion of the frequencies in the basis of the eigenvectors

$$\delta p_i^t = \alpha_1 \lambda_1^t v_{1i} + \alpha_2 \lambda_2^t v_{2i} + \cdots + \alpha_{2^B} \lambda_{2^B}^t v_{2^B i} \tag{14}$$

would lead to $\alpha_1 = 0$ for any initial condition. This observation, which is a consequence of the normalization of the frequencies, together with the fact that in the $2^B - 1$ relevant eigendirections the perturbations decrease, demonstrate that the equilibrium is stable.

We now calculate the Jacobian matrix

$$J_{i,j} = 2^{1-B} \sum_k (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma . \tag{15}$$

Although this sum looks similar to (1) the result is very different. For fixed i and j we sum in order of increasing $\gamma = d(j,k)$. In order to make the procedure clear we first work out an example with $B = 7$ and

$$\begin{aligned} i &= (0000000) \\ j &= (0000111). \end{aligned}$$

In this case $\alpha = (3 - \gamma + d(i,k))/2$.

For $\gamma = 0$, $k = j$ and $\alpha = d(i,j) = 3$.

For $\gamma = 1$ there are two basic possibilities for k: the first comprises three cases, (0000011); (0000101); (0000110), all with $d(i,k) = 2$ and $\alpha = 2$. The second comprises four cases, (0001111); (0010111); (0100111); (1000111), with $d(i,k) = 4$ and $\alpha = 3$.

For $\gamma = 2$ there are three possible configurations for k: the first has a single 1 in the last three positions, like (0000001), counting 3 cases, all with $d(i,k) = 1$ and $\alpha = 1$. The second has two 1's in the last three positions and one 1 in the first four positions, like (1000011), counting 12 cases, all with $d(i,k) = 3$ and $\alpha = 2$. Finally, the third configuration has two 1's in the first four positions and all last three positions are 1, like (0110111), 6 cases, with $d(i,k) = 5$ and $\alpha = 3$.

In general, for each γ there are $\gamma + 1$ configurations for $d(i,k)$, with

$$\begin{aligned} d(i,k) &= d(i,j) - (\gamma - 2k) \\ \alpha &= d(i,j) - \gamma + k \end{aligned}$$

for $k = 0, 1, \ldots, \gamma$. For each k the number of cases with $d(i,j)$ and α fixed is

$$\begin{pmatrix} d \\ \gamma - k \end{pmatrix} \begin{pmatrix} B - d \\ k \end{pmatrix}$$

where we abbreviate $d(i,j) = d$. Therefore,

$$J_{i,j} = 2^{1-B} \sum_{\gamma=0}^{B} \sum_{k=0}^{\gamma} \begin{pmatrix} d \\ \gamma - k \end{pmatrix} \begin{pmatrix} B - d \\ k \end{pmatrix} (1-\mu)^{B-\alpha-\gamma} \mu^{\alpha} \left(\frac{1}{2}\right)^{\gamma}. \tag{16}$$

For d fixed, the sum over k actually runs only from $\gamma - d$ to γ. It is convenient to define $j = \gamma - k$ and to sum over γ first:

$$J_{i,j} = 2^{1-B} (1-\mu)^{B-d} \mu^{d} \sum_{j=0}^{d} \begin{pmatrix} d \\ j \end{pmatrix} \sum_{\gamma=0}^{B} \begin{pmatrix} B - d \\ \gamma - j \end{pmatrix} \left(\frac{\mu}{1-\mu}\right)^{\gamma - j} \left(\frac{1}{2\mu}\right)^{\gamma}. \tag{17}$$

The sum over γ contributes only starting at $\gamma = j$ and we can simplify

$$\sum_{\gamma=j}^{B} \begin{pmatrix} B-d \\ \gamma-j \end{pmatrix} \left(\frac{\mu}{1-\mu}\right)^{\gamma-j} \left(\frac{1}{2\mu}\right)^{\gamma}$$

$$= \left(\frac{1-\mu}{\mu}\right)^{j} \sum_{\gamma=0}^{B-j} \begin{pmatrix} B-d \\ \gamma \end{pmatrix} \left(\frac{1}{2(1-\mu)}\right)^{\gamma+j}$$

$$= \left(\frac{1}{2\mu}\right)^{j} \sum_{\gamma=0}^{B-d} \begin{pmatrix} B-d \\ \gamma \end{pmatrix} \left(\frac{1}{2(1-\mu)}\right)^{\gamma} \tag{18}$$

$$= \left(\frac{1}{2\mu}\right)^{j} \left(1+\frac{1}{2(1-\mu)}\right)^{B-d} = \left(\frac{1}{2\mu}\right)^{j} \left(\frac{3-2\mu}{2(1-\mu)}\right)^{B-d}.$$

Substituting this result back into (17) we finally obtain

$$\begin{aligned} J_{i,j} &= 2^{1-2B+d}(3-2\mu)^{B-d}\mu^{d} \sum_{j=0}^{d} \begin{pmatrix} d \\ j \end{pmatrix} \left(\frac{1}{2\mu}\right)^{j} \\ &= 2^{1-2B+d}(3-2\mu)^{B-d}\mu^{d} \left(1+\frac{1}{2\mu}\right)^{d} \\ &= 2^{1-2B}(3-2\mu)^{B-d}(1+2\mu)^{d}. \end{aligned} \tag{19}$$

V. UNIFORM DISTRIBUTION FOR SIMILARITY-ESSENTIAL RECOMBINATION

Here we show that the uniform solution $p_i = 2^{-B}$ is also a solution of the equation

$$p_i^{t+1} = \sum_{j,k,\gamma \le G} p_j^t \, p_k^t \, c_\mu(j,k;i) - p_i^t(\Phi - 1). \tag{20}$$

with

$$\Phi = \sum_{j,k,\gamma \le G} p_j^t \, p_k^t \tag{21}$$

for arbitrary B and G.

Substituting $p_j = p_k = 2^{-B}$ in eq. (21) we find

$$\Phi = 1 - \sum_{j,k,\gamma > G} p_j \, p_k = 1 - 2^{-2B} \sum_{j,k,\gamma > G} 1. \tag{22}$$

For j fixed there is only one way to have $\gamma = B$, B different ways to have $\gamma = B - 1$, etc. Accordingly,

$$\Phi = 1 - 2^{-2B} \sum_j \sum_{k=G+1}^{B} \begin{pmatrix} B \\ k \end{pmatrix} = 1 - 2^{-B} \sum_{k=G+1}^{B} \begin{pmatrix} B \\ k \end{pmatrix}. \tag{23}$$

The first term of (20) computed at the uniform solution gives

$$\mathcal{N} \equiv \sum_{j,k,\gamma \leq G} p_j^t \, p_k^t \, c_\mu(j,k;i) = 2^{-2B} \sum_{j,k,\gamma \leq G} (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma. \tag{24}$$

This sum can be evaluated with the same method used in section I:

$$\begin{aligned} \mathcal{N} &= 2^{-2B} \sum_{k=0}^{G} \begin{pmatrix} B \\ k \end{pmatrix} \\ &= 2^{-2B} \left[2^B - \sum_{k=G+1}^{B} \begin{pmatrix} B \\ k \end{pmatrix} \right] \\ &= 2^{-2B} \left[2^B - (1-\Phi) 2^B \right] = 2^{-B} \Phi. \end{aligned} \tag{25}$$

Therefore, the right hand side of eq.(20) becomes $\mathcal{N} - 2^{-B}(\Phi - 1) = 2^{-B}$ and the uniform distribution indeed satisfies the equations.

VI. CRITICAL MUTATION RATES IN THE SIMILARITY-ESSENTIAL RECOMBINATION MODEL

The linearization of Eq.(20) has contributions coming from the first term $\mathcal{N}$ and from Φ. Setting $p_i = 2^{-B} + \delta p_i$ we obtain

$$\delta p_i^{t+1} = 2^{-B+1} \sum_{j,k,\gamma \leq G} \delta p_j^t \, (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma - \delta p_i (\Phi - 1) - 2^{-B} \delta \Phi \tag{26}$$

where

$$\delta \Phi = 2^{-B+1} \sum_{j,k,\gamma > G} \delta p_j^t = 2(1-\Phi) \sum_j \delta p_j^t. \tag{27}$$

However, since $\sum_j p_j = 1 = \sum_j (2^{-B} + \delta p_j^t) = 1 + \sum_j \delta p_j^t$, we find that $\sum_j \delta p_j^t = 0$. Therefore,

$$\delta p_i^{t+1} = \sum_j J_{i,j} \, \delta p_j^t \tag{28}$$

where

$$J_{i,j} = 2^{-B+1} \sum_j \left[\sum_{k,\gamma \leq G} (1-\mu)^{B-\alpha-\gamma} \mu^\alpha \left(\frac{1}{2}\right)^\gamma - (\Phi - 1) \delta_{ij} \right]. \tag{29}$$

The non-diagonal term can be rewritten using eq.(17) and we find

$$\frac{(1-\mu)^{B-d}\mu^d}{2^{B-1}} \sum_{\gamma=0}^{G} \sum_{j=0}^{\gamma} \begin{pmatrix} d \\ j \end{pmatrix} \begin{pmatrix} B-d \\ \gamma-j \end{pmatrix} \left(\frac{\mu}{1-\mu}\right)^{\gamma-j} \left(\frac{1}{2\mu}\right)^{\gamma}, \tag{30}$$

where we abbreviate $d(i,j) = d$. Inverting the sums this can be further rewritten as

$$\sum_{\gamma=0}^{G} \sum_{j=0}^{\gamma} \rightarrow \sum_{j=0}^{j_{max}} \sum_{\gamma=j}^{\gamma_{max}} \tag{31}$$

where the $j_{max} = min[d, G]$ and $\gamma_{max} = min[j + B - d, G]$ are fixed by the binomial terms. The Jacobian matrix depends only on $d(i,j)$ and is given by

$$J_{i,j} = f(d(i,j)) - (\Phi - 1)\delta_{ij} \tag{32}$$

with

$$f(d) \ = \frac{(1-\mu)^{B-d}\mu^d}{2^{B-1}} \sum_{j=0}^{j_{max}} \sum_{\gamma=j}^{\gamma_{max}} \begin{pmatrix} d \\ j \end{pmatrix} \begin{pmatrix} B-d \\ \gamma-j \end{pmatrix} \left(\frac{1-\mu}{\mu}\right)^{j} \left[\frac{1}{2(1-\mu)}\right]^{\gamma}. \tag{33}$$

We find that the uniform solution always becomes unstable for sufficiently small mutation rates, $\mu < \mu_c$ where μ_c depends on B and G. By explicit computation of the eigenvalues of the Jacobian matrix for small values of B and G we can derive expressions for $\mu_c(B,G)$ (see section III.D of the paper). For larger values of G the expressions become more complicated and a more practical method of computation is given below.

As illustrations we consider the simple case $B = 2$ and two values of G. The uniform solution is $p_{00} = p_{01} = p_{10} = p_{11} = 1/4$.

– $G = 1$. The three independent entries of the 4×4 Jacobian matrix are given by

$$\begin{aligned} f(0) \ &= \frac{1}{2}(2-\mu)(1-\mu) \\ f(1) \ &= \frac{1}{4}(1 + 2\mu - 2\mu^2) \\ f(2) \ &= \frac{1}{2}\mu(1+\mu) \end{aligned} \tag{34}$$

and $\Phi = 3/4$. The Jacobian matrix itself is

$$J = \begin{pmatrix} f(0)+1/4 & f(1) & f(1) & f(2) \\ f(1) & f(0)+1/4 & f(2) & f(1) \\ f(1) & f(2) & f(0)+1/4 & f(1) \\ f(2) & f(1) & f(1) & f(0)+1/4 \end{pmatrix} . \tag{35}$$

The eigenvalues are $\lambda_1 = 7/4$ (corresponding to the vector $(1,1,1,1)$), $\lambda_2 = \lambda_3 = (5-8\mu)/4$ and $\lambda_4 = (3-8\mu+8\mu^2)/4$. Although λ_4 is always bounded between 0 and 1, λ_2 and λ_3 become larger than 1 at $\mu_c = 1/8$, when the uniform solution becomes unstable. The two corresponding eigenvectors are $v_2 = (0,1,-1,0)$ and $v_3 = (1,0,0,-1)$.

– $G = 0$. In this case

$$\begin{aligned} f(0) &= \frac{1}{2}(1-\mu)^2 \\ f(1) &= \frac{1}{2}\mu(1-\mu) \, . \\ f(2) &= \frac{1}{2}\mu^2 \end{aligned} \tag{36}$$

and $\Phi = 1/4$. The eigenvalues are $\lambda_1 = 5/4$ (corresponding to the vector $(1,1,1)$), $\lambda_2 = \lambda_3 = (5-4\mu)/4$ and $\lambda_4 = (5-8\mu+8\mu^2)/4$. Now λ_2 and λ_3 become larger than 1 at $\mu_{\mathrm{crit}} = 1/4$. The two corresponding eigenvectors are once again $v_2 = (0,1,-1,0)$ and $v_3 = (1,0,0,-1)$.

This structure generalizes for any B and G: the eigenvalue that can become larger than 1 has the form $\alpha(1-2\mu)$, where α is a numerical factor, and one of its corresponding eigenvector has four blocks of size $B/4$ each, $v = (0,0,..,0,1,1...1,-1,-1,...-1,0,0...,0)$. Imposing $Jv = v$ we find that μ_c corresponds to the solution of the equation

$$\sum_{d=0}^{B} c(d,B) f(d,\mu) = \Phi. \tag{37}$$

where the coefficients $c(j,B)$ satisfy the recursion relation

$$c(j,B) = c(j,B-1) + c(j-1,B-1) \tag{38}$$

with $c(0,B) = 1$, $c(1,1) = -1$ and $c(j,B) = 0$ for $j > B$ and $B \geq 2$.

The linearity of the eigenvalue from which we extract μ_c means that the terms of order greater or equal than 2 in μ_c should vanish. Accordingly, we can write:

$$T_1 + \mu_c T_2 = \Phi \tag{39}$$

where

$$T_1 = \sum_{d=0}^{B} c(d,B) \lim_{\mu \to 0} f(d,\mu) \tag{40}$$

and

$$T_2 = \sum_{d=0}^{B} c(d,B) \lim_{\mu \to 0} \frac{\partial f}{\partial \mu}(d,\mu) \tag{41}$$

which leads to the following closed formula for μ_c

$$\mu_c = \frac{\Phi - \sum_{d=0}^{B} c(d,B) \lim_{\mu \to 0} f(d,\mu)}{\sum_{d=0}^{B} c(d,B) \lim_{\mu \to 0} \frac{\partial f}{\partial \mu}(d,\mu)}. \tag{42}$$