An evolution model that satisfies detailed balance

Jüri Lember
University of Tartu, Estonia

Suppose we have a finite set $\mathcal{X} = \{1, \ldots, K\}$ of possible genomes, and $\xi_1, \xi_2, \ldots$ denotes an exchangeable $\mathcal{X}$-valued stochastic process (the breeding process). For any population of n genomes $x_1, \ldots, x_n$, we breed an $n+1$'th genome by sampling x_{n+1} from the distribution $P(\xi_{n+1} = \cdot | \xi_1 = x_1, \ldots, \xi_n = x_n)$. To every genotype $x \in \mathcal{X}$, a fitness $w(x) > 0$ is assigned and after $n+1$'th genome/individual is sampled, one of the $n+1$ individuals is removed in a fitness-biased way so that that less fit individuals are less likely to be discarded. There are several ways (selection schemes) for doing that and as a result one ends with a Markov chain with state space consisting of n-tuples of genotypes. We introduce several selection schemes under which the Markov chain satisfies detailed balance equality with stationary measure being

$$\underbrace{P_n(x_1, \ldots, x_n)}_{\text{stationary distribution}} = \frac{1}{Z_n} \underbrace{P_\xi(x_1, \ldots, x_n)}_{\text{breeding term}} \underbrace{w(x_1) \cdots w(x_n)}_{\text{fitness term}}.$$

When the breeding process ξ is Dirichlet-categorical process (or Polya urn), then the above-described model contains several versions of Moran models as special cases. We are mainly interested in the limit of P_n as population size n grows. We allow fitness w_n depend on populations size and take $w_n(x) = \exp[-\phi(x)/n^\lambda]$. We show that when $\lambda \in [0, 1)$, then only the fittest genotype survives, when $\lambda = 1$, then the proportions of genotypes converge to a nondegenerate measure, and when $\lambda > 1$, then the influence of fitness vanishes.

We also consider the case when breeding process is Dirichlet-categorical process with the Dirichlet measure $\mathrm{Dir}(n^{1-\lambda}\alpha_1, \ldots n^{1-\lambda}\alpha_K)$. Again, the phase transition results are proven. For some special cases, some of the obtained limits can be found in the literature, typically obtained via diffusion approximation. Our derivation is based on fairly simple arguments without any approximation. The mathematics, however, becomes much more involved when the set of genotypes is infinite.

This is a joint work with C. Watkins (RHUL).