Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 1681–1709

stochastic processes and their applications

www.elsevier.com/locate/spa

The distribution of the quasispecies for a Moran model on the sharp peak landscape

Raphaël Cerf, Joseba Dalmau*

Université Paris Sud, France
DMA, École Normale Supérieure, France

Received 17 September 2014; received in revised form 1 December 2015; accepted 2 December 2015
Available online 11 December 2015

Abstract

We consider the Moran model on the sharp peak landscape, in the asymptotic regime studied in Cerf (2015), where a quasispecies is formed. We find explicitly the distribution of this quasispecies.

Keywords: Quasispecies; Eigen's model; Moran model

1. Introduction

In his paper [10], Eigen introduced the model of quasispecies to describe the evolution of a population of macromolecules which is subject to two main forces: mutation and selection. The model was developed further in a series of papers by Eigen and Schuster [12–14], and analysed in great detail by Eigen, McCaskill and Schuster in [11]. A major conclusion is that this kind of evolutionary process, rather than selecting a single dominant species, is more likely to select a master sequence (the macromolecule with the highest fitness) along with a cloud of mutants that closely resemble the master sequence. Hence the name quasispecies. One other major discovery that Eigen made was the existence of an error threshold allowing a quasispecies to form: if the mutation rate exceeds the error threshold, then the population evolves towards a totally random state, whereas if the mutation rate is below the error threshold, a quasispecies can be formed.

Even if Eigen's original goal was to explain the behaviour of a population of macromolecules, the theory of quasispecies rapidly extended to other areas of biology. In particular, experimental studies support the validity of the model in virology [9]. Some RNA viruses are known to have very high mutation rates, like the HIV virus, and this is a factor of resistance against conventional drugs. A promising strategy to combat this kind of viruses consists in developing mutagenic

* Corresponding author.
E-mail address: dalmau@dma.ens.fr (J. Dalmau).

http://dx.doi.org/10.1016/j.spa.2015.12.002
0304-4149/

drugs that would increase the mutation rate beyond the error threshold, in order to induce an error catastrophe [2,24]. This strategy has successfully been applied to several types of RNA viruses [5]. Moreover, several similarities have been observed between the evolution of cancer cell populations and RNA viruses, in particular, the possibility of inducing an error catastrophe [23].

Two important features of Eigen's model are its deterministic nature (the model is based on a system of differential equations derived from certain chemical and physical laws) and the fact that the population is considered to be infinite. When dealing with simple macromolecules, these assumptions are quite natural. Nevertheless, they become unrealistic if we want to apply this model to population genetics, and they are two of the major drawbacks when applying it to virus populations, as pointed out by Wilke [25]. On one hand, we have to take into account the stochastic nature of the evolution of a finite population. The higher the complexity of the individuals, the harder it is to explain the replication and mutation schemes via chemical reactions. This fact, together with the widely recognised role of randomness in evolutionary processes strongly suggest a stochastic approach to the matter. On the other hand, when dealing with populations of complex individuals, the amount of possible genotypes largely exceeds the size of the population. Therefore, if we want to use Eigen's model in population genetics, a finite and stochastic version of the model is called for.

The interest of a finite stochastic counterpart to Eigen's model is not new. Eigen, McCaskill and Schuster already emphasise the importance of developing such a model [11], so does Wilke in the more recent paper [25]. Several researchers have pursued this task. Demetrius, Schuster and Sigmund [7] introduce stochasticity into Eigen's model using branching processes. McCaskill [17] also develops a stochastic version of Eigen's model. Nowak and Schuster [20] use birth and death Markov processes to give a finite stochastic version of Eigen's model on the sharp peak landscape. Alves and Fontanari [1] study the dependence of the error threshold on the population size for the sharp peak replication landscape. Saakian, Deem and Hu [22] compute the variance of the mean fitness in a finite population model in order to control how it approximates the infinite population model. Deem, Muñoz and Park [21] use a field theoretic representation in order to derive analytical results. Other recent papers introduce finite stochastic models that approach Eigen's model asymptotically when the population size goes to ∞, like Musso [19] or Dixit, Srivastava, Vishnoi [8].

In [3], Cerf studies a population of size m of chromosomes of length ℓ over an alphabet $\mathcal{A}$ of cardinality κ, which evolves according to a Moran model [18]. The mutation probability per locus is q. Only the sharp peak landscape is considered: the master sequence, which we denote by w^*, replicates with rate $\sigma > 1$, while all the other sequences replicate with rate 1. In the asymptotic regime where

$$\ell \to +\infty, \qquad m \to +\infty, \qquad q \to 0,$$
$$\ell q \to a, \qquad \frac{m}{\ell} \to \alpha,$$

a critical curve is obtained in the parameter space (a, α), which is given by $\alpha\phi(a) = \ln\kappa$. If $\alpha\phi(a) < \ln\kappa$, then the population is totally random, i.e., the fraction of the master sequence in a population at equilibrium converges to 0. On the contrary, if $\alpha\phi(a) > \ln\kappa$, then a quasispecies is formed, i.e., at equilibrium, the population contains a positive fraction of the master sequence, which in the asymptotic regime presented above converges to $(\sigma e^{-a} - 1)/(\sigma - 1)$.

The aim of our article is to obtain the whole distribution of the quasispecies. As it is customary with this kind of models, we introduce Hamming classes with respect to the master sequence in the space $\mathcal{A}^\ell$ of sequences of length ℓ. We say that a chromosome $u \in \mathcal{A}^\ell$ belongs to the class