Mathematical Biosciences 285 (2017) 61–67

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

A generalization of Kingman's model of selection and mutation and the Lenski experiment

Linglong Yuan[1]

Institut für Mathematik, Johannes-Gutenberg-Universität Mainz, Staudingerweg 9, 55099 Mainz, Germany

ARTICLE INFO

Article history:
Received 19 September 2016
Revised 18 December 2016
Accepted 20 December 2016
Available online 23 December 2016

92D19
37N25
60F05
92D15

Keywords:
Selection-mutation balance
House of cards
Type distribution
Fitness function
Lenski experiment
Escherichia coli

ABSTRACT

Kingman's model of selection and mutation studies the limit type value distribution in an asexual population of discrete generations and infinite size undergoing selection and mutation. This paper generalizes the model to analyze the long-term evolution of Escherichia. coli in Lenski experiment. Weak assumptions for fitness functions are proposed and the mutation mechanism is the same as in Kingman's model. General macroscopic epistasis are designable through fitness functions. Convergence to the unique limit type distribution is obtained.

1. Introduction

Evolutionary forces in a population vary from macroscopic scale to microscopic scale including random environment, migration, natural selection, macroscopic epistasis (or individual interaction), microscopic epistasis, and linkage and dominance, clonal interference, mutation, genetic drift, recombination, and so on. Recently, mathematicians are interested in incorporating as many factors as possible in an evolutionary model, either deterministic or stochastic, to understand the contribution of each factor and to see which state the model can reach in the limit (see for example [4,5,15] among numerous works). However one would expect a high level of complexity of modelling and analysis when many factors enter into play.

Kingman [10] suggested that one can regard an equilibrium of the evolutionary model as existing because of two preponderant factors, other phenomena causing perturbations of the equilibrium. The pair of factors in his model are selection and mutation. This particular case had also been the subject of study of Moran [12–14] almost at the same time.

More specifically, Kingman [10] proposed a one-locus, discrete generation model under selection and mutation with an infinite number of possible alleles which have continuous effects on a quantitative type. The continuum-of-alleles models were introduced by Crow and Kimura [6] and Kimura [9] and are used frequently in quantitative genetics, since types usually have a polygenetic basis.

Kingman's idea can be applied to model the Lenski experiment which investigates the long term evolution of E.coli in the laboratory. Indeed, the application goes to various evolutionary models and one major parameter is how selection influences the population. That generates many variants of Kingman's model and a general treatment is required.

The paper aims to establish a general model which covers the Kingman's setting and can be applied to Lenski experiment. In Section 2, we show briefly the Kingman's model and the main observations in Lenski experiment. This section is the motivation of the paper, but the reader can skip it for the first reading since we will come back for applications. In Section 3, we introduce a general setting with 3 assumptions on the fitness function. We give the main results for the general model when some or all assumptions hold. Section 4 is devoted to proofs and in Section 5, we

E-mail address: yuanlinglongcn@gmail.com
URL: http://linglongyuan.weebly.com

[1] Linglong Yuan gratefully acknowledges support by DFG priority programme SPP 1590 *Probabilistic structures in evolution*.

http://dx.doi.org/10.1016/j.mbs.2016.12.007
0025-5564/

show the applications to Kingman's model and Lenski experiment. Section 6 summarizes the main contribution of the paper and discusses the comparison of our model with other works, especially with [8] on Lenski experiment.

2. Kingman's model and Lenski experiment

2.1. Kingman's model

The model considers an effectively infinite population that reproduces asexually and has discrete generations. It studies a specific type and the selection influences the population through the fitness which is the offspring size and depends (possibly not only) on its type value x, a real number in the space $\mathcal{M} := [0, M] \subset \mathbb{R}_0^+$ where M is a positive real number. Let $\mathbb{P}(\mathcal{M})$ denote the set of probability measures on $\mathcal{M}$. For any $u \in \mathbb{P}(\mathcal{M})$, let m_u denote the upper limit of the support of u, i.e., $m_u = \sup\{x : x \in \mathcal{M}, u([x-\varepsilon, x]) > 0, \forall 0 < \varepsilon < x\}$. So m_u is the largest type value an individual can take in a population with type distribution u.

Assume that each individual per generation mutates independently with probability β $(0 < \beta < 1)$ and the mutant type distribution is the probability measure q on $\mathcal{M}$, independent of parent's type. Kingman [10,11] argued that the tendency for most mutations to be deleterious might be reflected in a model in which the gene after mutation is independent of that before, the mutation having destroyed the biochemical "house of card" built up by evolution.

The fitness function in this model is $x \mapsto x, x \in \mathcal{M}$. Let $(p_i)_{i \geq 0}$ denote the sequence of type distributions of generations i on $\mathcal{M}$ with p_0 given as a parameter; Then $(p_i)_{i \geq 0}$ are defined recursively:

$$p_i(dx) = (1-\beta)\frac{xp_{i-1}(dx)}{\int xp_{i-1}(dx)} + \beta q(dx), i \geq 1. \tag{1}$$

In particular, we set $p_i(dx) = (1-\beta)p_{i-1}(dx) + \beta q(dx)$, if $p_{i-1} = \delta_0$, the Dirac measure at 0.

Remark 1. Due to the expression of (1), it is clear that $m_{p_i} \leq \max\{m_{p_0}, m_q\}$ for any $i \geq 0$. So letting $M = \max\{m_{p_0}, m_q\}$ and $\mathcal{M} = [0, \max\{m_{p_0}, m_q\}]$ does not change any p_i. Since $m_q \leq m_{p_1}$, one can assume $m_q \leq m_{p_0}$, otherwise we take p_1 as p_0. For convenience, we introduce:

Convention (*):

$$m_q \leq m_{p_0}, M = m_{p_0} \text{ and } \mathcal{M} = [0, m_{p_0}].$$

If a sequence of measures (not necessarily probability measures) $(h_i)_{i \geq 0}$ converges in total variation sense to a measure h, that is, the total variation of $h_i - h$ tending to 0, then for abbreviation, we say $(h_i)_{i \geq 0}$ converges strongly to h.

Kingman specifically takes $M = 1$ in his model. Based on the value of $\int \frac{q(dx)}{1-x}$, Kingman [10] proved that:

Theorem 1. *(Kingman) Case 1:* $\int \frac{q(dx)}{1-x} > \beta^{-1}$. *Then* $(p_i)_{i \geq 0}$ *converges strongly to*

$$p^*(dx) = \frac{\beta s q(dx)}{s - (1-\beta)x},$$

with s *being the unique solution of* $\int \frac{\beta s q(dx)}{s-(1-\beta)x} = 1$.

Case 2: $\int \frac{q(dx)}{1-x} \leq \beta^{-1}$. *Then* $(p_i)_{i \geq 0}$ *converges weakly to*

$$p^*(dx) = \frac{\beta q(dx)}{1-x} + (1 - \int \frac{\beta q(dy)}{1-y})\delta_1(dx),$$

here $\delta_1(dx)$ *is the Dirac measure at 1.*

Therefore the sequence $(p_i)_{i \geq 0}$ converges at least in the weak sense to a limit distribution p^* which depends only on q and β, regardless of the specific form of p_0. Biologically, it can be seen as a stability property of the population.

Next we introduce the Lenski experiment and use an iteration similar to (1) to model the evolution of E.coli.

2.2. Lenski experiment and modelling

The Lenski experiment is a long-term evolution experiment with E.coli, founded by Richard E. Lenski in 1988 in the laboratory. The experiment is decomposed into *daily cycles*. Every day starts by sampling approximately $5 \cdot 10^6$ bacteria from those available in the medium that was used the previous day. This sample is then transferred to a new glucose-limited minimal medium and reproduce (asexually) until the medium is deployed, i.e., when there is no more glucose available. Around $5 \cdot 10^8$ cells are present at the end of each day. So the size grows by approximatively 100 times from the beginning of a day to the end and a sample of percentage around 1% will be chosen for the next day. The closely 30-year ongoing experiment has run more than 60,000 generations. We refer to [8] for a more detailed presentation and references therein.

There are 12 populations founded from a common ancestor. Samples, called by Lenski "fossil record", are frozen every 500 generations. Once bacterium is frozen, we consider it stopping biological activities inside the body, which is how the name "fossil record" makes sense. The records are regarded as stocked information of evolutionary trajectories of populations.

They define the fitness as the dimensionless ratio of the competitors' realized reproduction rates. Basically, we let two populations of the same number of individuals, one of the founder ancestors and one of evolved strain, to be together in a medium at the beginning of a day. The fitness of the evolved strain is the ratio of the (exponential) reproduction rate of the strain observed at the end of the day and the reproduction rate of the ancestor strain. So in this definition, fitness is a *relative* quantity that measures the reproduction rate of the whole population. However mathematically, one can directly model the natural (non-relative) reproduction rate of each bacterium. To unify notations, we shall consider the natural reproduction rate as the type and our fitness, different from that of Lenski, is the offspring size in the next generation.

Wiser et al. [16] showed that the (relative) reproduction rate increases but decelerates. They compared the hyperbolic and sublinear power law increasing models, the former having a bound and the latter none. It turns out that the hyperbolic model fits to the first 10,000 generations, but for a long term about 50,000 generations, the sublinear power law model is more significant.

The unboundedness of the sublinear power law curve can be explained by the fact that highly beneficial mutations happen rarely but consistently and with probability 1 some of them fixate the population, although after a probably very long time. Therefore, on a very long period of time, one can consider that there is a bound (or even a pattern) for the reproduction rates of new mutants.

More specifically, let p_0 be the initial type distribution (or reproduction rate distribution) and q the mutant type distribution such that $0 \leq m_q \leq m_{p_0} < \infty$. Let $M, \mathcal{M}$ be defined from p_0, q by convention (*). Let the population grow exponentially according to the reproduction rate of each individual until the total amount reaches the capacity $\gamma (\gamma > 1)$ (assuming the initial amount 1). In Lenski experiment, $\gamma \approx 100$. We then sample $1/\gamma$ proportion of the population at the end of day 0 to constitute the population at the beginning of day 1. However, we have new mutants arriving along the whole day. To combine the mutation and selection together, we assume that a $\beta (0 < \beta < 1)$ proportion of the sample consists of mutation population with type distribution q and the rest $1-\beta$ proportion stays unchanged. For any $u \in \mathbb{P}(\mathcal{M})$, let t_u be the unique solution of $\int e^{t_u x} u(dx) = \gamma$. The type distribution p_i at