

# KINGMAN'S MODEL WITH RANDOM MUTATION PROBABILITIES: CONVERGENCE AND CONDENSATION I

LINGLONG YUAN,\* *University of Liverpool and Xi'an Jiaotong-Liverpool University*

## Abstract

For a one-locus haploid infinite population with discrete generations, the celebrated model of Kingman describes the evolution of fitness distributions under the competition of selection and mutation, with a constant mutation probability. This paper generalises Kingman's model by using independent and identically distributed random mutation probabilities, to reflect the influence of a random environment. The weak convergence of fitness distributions to the globally stable equilibrium is proved. Condensation occurs when almost surely a positive proportion of the population travels to and condenses at the largest fitness value. Condensation may occur when selection is favoured over mutation. A criterion for the occurrence of condensation is given.

*Keywords:* Population dynamics; mutation–selection balance; house of cards; fitness distribution; size-biased distribution; distributional equation

2020 Mathematics Subject Classification: Primary 60F05

Secondary 60G10; 92D15; 92D25

## 1. Motivation and background

Various biological forces interact with each other and jointly drive the evolution of populations. One important competing pair consists of selection and mutation. As early as 1937, Haldane [14] put forward the concept of mutation–selection balance. The mathematical foundation of this subject was established by Crow and Kimura [7], Ewens [11], and Kingman [20]. For more details on this topic, we refer to Bürger [5, 6].

A simple setting is to consider a one-locus haploid infinite population with discrete generations under selection and mutation. The locus is assumed to have infinitely many possible alleles, which have continuous effects on a quantitative type. The continuum-of-alleles models were introduced by Crow and Kimura [7] and Kimura [17] and are used frequently in quantitative genetics.

Kingman [18] suggested that the tendency that most mutations are deleterious could be explained by the assumption of the independence of the gene before and after mutation. The paper [19], which proposed Kingman's famous one-locus model, described this feature in terms of a 'house of cards'; that is, a mutation destroys the biochemical house of cards built up by evolution. In the one-locus model, a population is characterised by its type distribution, which is a probability measure on  $[0,1]$ , and any  $x \in [0, 1]$  is a type value. In Kingman's setting, an

---

Received 26 January 2020; revision received 30 May 2021.

\* Postal address: University of Liverpool, Department of Mathematical Sciences, Peach Street, L69 7ZL, Liverpool, UK. Email address: [yuanlinglongcn@gmail.com](mailto:yuanlinglongcn@gmail.com)

© The Author(s), 2022. Published by Cambridge University Press on behalf of Applied Probability Trust.

individual with a larger type value is fitter, which means more productive. So the type value can also be referred to as a fitness value. Kingman's model can be seen as the limit of a finite-population model; see [13].

Bürger [4] generalised the selection mechanism by allowing the gene after mutation to depend on that before and proved convergence in total variation. The genetic variation of the equilibrium distribution was computed and discussed. In [22] the author of the present paper proposed a more general selection mechanism which can model general macroscopic epistasis, with the other settings the same as in Kingman's model. This model was applied to the modelling of the Lenski experiment (see [12] for a description of the experiment).

There also exist models on the balance of selection and mutation in the setting of continuous generations. Bürger [3] provided an exact mathematical analysis of Kimura's continuum-of-alleles model, focusing on the equilibrium genetic variation. Steinsaltz *et al.* [21] proposed a multi-locus model using a differential equation to study the ageing effect. Later on, recombination was incorporated into the model [10]. The model of Betz *et al.* [2] generalised a continuous-time version of Kingman's model and other models arising from physics.

However, to the best of the author's knowledge, Kingman's model has never been generalised to a random version. In this paper we will assume that the mutation probabilities of all generations form an independent and identically distributed (i.i.d.) sequence. Biologically, we think of a stable random environment such that the mutation probabilities vary with time but are independently sampled from the same distribution.

In Kingman's model, condensation occurs if a certain positive proportion of the population travels to and condenses at the largest fitness value. This is due to the dominance of selection over mutation. In the random model proposed in this paper, we also consider the convergence of (random) fitness distributions to the equilibrium and the condensation phenomenon. Moreover, Kingman's model has been revisited recently in terms of the travelling wave of mass to the largest fitness value [8]. The random model provides another example for consideration in this direction.

## 2. Models

### 2.1. Kingman's model with time-varying mutation probabilities

Consider a haploid population of infinite size and discrete generations under the competition of selection and mutation. We use a sequence of probability measures  $(P_n) = (P_n)_{n \geq 0}$  on  $[0,1]$  to describe the distribution of fitness values in the  $n$ th generation. We can assume, more generally, that the probability measures are supported on a finite interval, not necessarily  $[0,1]$ . But since only fitness ratios will be relevant (see [19] or [22] for a more explicit explanation), we adopt the setting of  $[0,1]$ , which was used by Kingman [19], and which is equivalent to general finite supports.

Individuals in the  $n$ th generation are children of the  $(n-1)$ th generation. First of all, the fitness distribution of children is initially  $P_{n-1}$  (an exact copy from the parents). Then selection takes effect, updating the fitness distribution from  $P_{n-1}$  to the size-biased distribution

$$\frac{xP_{n-1}(dx)}{\int yP_{n-1}(dy)}$$

Here and henceforth, we use  $\int$  to denote  $\int_0^1$ . Basically, the new population is re-sampled from the existing population by using their fitness as a selective criterion. Next, each individual mutates independently with the same mutation probability, which we denote by  $b_n$ ,

taking values in  $[0, 1]$ . Each mutant has fitness value sampled independently from a common mutant distribution, which we denote by  $Q$ , a probability measure on  $[0,1]$ . Then the resulting distribution is the distribution of the  $n$ th generation:

$$P_n(dx) = (1 - b_n) \frac{xP_{n-1}(dx)}{\int yP_{n-1}(dy)} + b_nQ(dx). \tag{1}$$

The reason we exclude the case that  $b_n$  equals 1 is that in this situation we have  $P_n = Q$ ; that is, we have completely lost the accumulated evolutionary changes. This is not interesting either biologically or mathematically.

Expanding (1), we can also obtain

$$P_n(dx) = \left( \prod_{l=0}^{n-1} \frac{1 - b_{l+1}}{\int yP_l(dy)} \right) x^n P_0(dx) + \sum_{j=1}^n \left( \prod_{l=j}^{n-1} \frac{1 - b_{l+1}}{\int yP_l(dy)} \right) b_j m_{n-j} Q^{n-j}(dx), \tag{2}$$

where

$$Q^k(dx) := \frac{x^k Q(dx)}{\int y^k Q(dy)}, \quad m_k := \int x^k Q(dx), \quad \forall k \geq 0.$$

In particular, if  $Q = \delta_0$ , the Dirac measure on  $\{0\}$ , then  $Q^k = \delta_0$  for any  $k \geq 0$ .

When all the  $b_n$  are equal to the same number  $b \in [0, 1]$ , this is the model introduced by Kingman [19]. In the general setting we allow the mutation probabilities to be different. We call it *Kingman's model with time-varying mutation probabilities*, or *the general model* for short. We introduce a few more pieces of notation. Let  $M$  be the space of (nonnegative) Borel measures on  $[0,1]$  and  $M_1$  the subspace of  $M$  consisting of probability measures. Let  $M, M_1$  be endowed with the topology of weak convergence. We use  $\xrightarrow{d}$  to denote weak convergence. We say a sequence of measures  $(u_n)$  converges in total variation to a measure  $u$ , and write  $u_n \xrightarrow{TV} u$ , if the total variation, defined as  $\sup_B |u_n(B) - u(B)|$  where the supremum is taken over all Borel sets, converges to 0.

For any  $u \in M_1$ , define

$$S_u := \sup\{x:u[x, 1] > 0\}. \tag{3}$$

We interpret  $S_u$  as the *largest fitness value* in a population of distribution  $u$ . Define  $h := S_{P_0}$ . It is not difficult to see that  $S_{P_n} = \max\{S_{P_0}, S_Q\}$  for any  $n \geq 1$ . Since we are interested in asymptotics, it is thus without loss of generality to assume that  $h \geq S_Q$ . Therefore  $S_Q \leq h \leq 1$ .

Note that the general model has parameters  $(b_n)_{n \geq 1}, Q, P_0, h$ . Kingman's model shares the same parameters, but with the  $b_n$  all equal to  $b$ . We call  $(P_n)$  *the forward sequence* or just *the sequence*. Although  $h$  is determined by  $P_0$ , we still consider  $h$  as a parameter, as it will become clear later that for Kingman's model and the random model considered in this paper, the limit of  $(P_n)$  depends on  $P_0$  only through  $h$ . This is the property of so-called global stability.

### 2.2. Convergence and condensation in Kingman's model

Kingman [19] proved the convergence of  $(P_n)$  when all mutation probabilities are equal, i.e.  $b_n = b$  for all  $n \geq 1$ .

**Theorem 1.** (Kingman's theorem, [19].) *1. If  $\int \frac{Q(dx)}{1-x/h} \geq b^{-1}$ , then  $(P_n)$  converges in total variation to*

$$\mathcal{K}(dx) = \frac{b\theta_b Q(dx)}{\theta_b - (1-b)x},$$

with  $\theta_b$ , as a function of  $b$ , being the unique solution of

$$\int \frac{b\theta_b Q(dx)}{\theta_b - (1-b)x} = 1. \tag{4}$$

2. If  $\int \frac{Q(dx)}{1-x/h} < b^{-1}$ , then  $(P_n)$  converges weakly to

$$\mathcal{K}(dx) = \frac{bQ(dx)}{1-x/h} + \left(1 - \int \frac{bQ(dy)}{1-y/h}\right) \delta_h(dx).$$

Note that  $\mathcal{K}$  is uniquely determined by  $b, Q, h$ , but not the choice of  $P_0$ . In this sense  $\mathcal{K}$  is a globally stable equilibrium. For simplicity, for any measure, say  $\mu$ , its mass on a point  $x$  is denoted by  $\mu(x)$  instead of  $\mu(\{x\})$ . Then we say there is condensation at  $h$  in Kingman’s model if  $Q(h) = 0$  but  $\mathcal{K}(h) > 0$ . We call  $\mathcal{K}(h)$  the condensate size if  $Q(h) = 0$ . In Case 1 above, there is no condensation. The condition

$$\int \frac{Q(dx)}{1-x/h} \geq b^{-1}$$

is satisfied only if  $b$  is big and/or  $Q$  is fit (i.e., having more mass on larger values). It means mutation is stronger than selection, so that the limit does not depend on  $P_0$  at all.

In Case 2, the condition

$$\int \frac{Q(dx)}{1-x/h} < b^{-1}$$

implies  $Q(h) = 0$ , but we have that  $\mathcal{K}(h) > 0$ . So there is condensation. In contrast to the first case, selection is favoured over mutation, so that the limit depends on  $P_0$  through  $h$ . If  $P_0(h) = 0$  (implying  $S_{P_n} = h$  and  $P_n(h) = 0$  for any  $n$ ), a certain amount of mass

$$\mathcal{K}(h) = \left(1 - \int \frac{bQ(dy)}{1-y/h}\right)$$

travels to the largest fitness value  $h$ , by the force of selection.

Next we introduce the random model, which is the main object of study in this paper.

**2.3. Kingman’s model with random mutation probabilities**

Let  $(\beta_n)_{n \geq 0}$  be an i.i.d. sequence of random variables in the common probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , taking values in  $[0,1)$  with common distribution  $\mathcal{L} \in M_1$  supported on  $[0,1)$ . *Kingman’s model with random mutation probabilities* or simply *the random model* is defined by the following dynamical system:

$$P_n(dx) = (1 - \beta_n) \frac{xP_{n-1}(dx)}{\int yP_{n-1}(dy)} + \beta_n Q(dx), \quad n \geq 1. \tag{5}$$

The random model has parameters  $(\beta_n), Q, P_0, h$ . It is a randomisation of Kingman’s model, as we can set each  $\beta_n$  to equal  $b$  with probability 1.

We are interested in the convergence of  $(P_n)$  to the equilibrium and the phenomenon of condensation. Since we are dealing with random probability measures, i.e., random elements of  $M_1$ , let us recall the definition of weak convergence in this context. Random (probability)

measures  $(\mu_n)$  supported on  $[0,1]$  converge weakly to a limit  $\mu$  if and only if for any continuous function  $f$  on  $[0,1]$  we have

$$\int f(x)\mu_n(dx) \xrightarrow{d} \int f(x)\mu(dx).$$

We refer to [16] for a reference on random measures. The definition of weak convergence for random measures stated in the follow-up paper [23] is incorrect. But this does not affect anything there as the weak convergence results are all proved in this paper.

As the sequence  $(P_n)$  is completely determined by  $(\beta_n)$ ,  $Q$ ,  $P_0$ , and  $h$ , the only randomness arises from  $(\beta_n)$ . In the terminology of statistical physics, the weak limit of  $(P_n)$  is an *annealed* limit, which is obtained given the law of  $(\beta_n)$ . A *quenched* limit, which is obtained by conditioning on  $(\beta_n)$ , does not exist unless  $P_0 = Q = \delta_0$ . A simple reason for nonexistence is that  $P_n$  contains  $\beta_n Q$ , which fluctuates persistently as  $(\beta_n)$  is i.i.d. However in Section 4.3 we will see that it is possible to obtain a quenched limit if the evolution is seen backwards.

For the particular case that  $Q = \delta_0$ , we have

$$P_n(dx) = (1 - \beta_n) \frac{x^n P_0(dx)}{\int y^n P_0(dy)} + \beta_n \delta_0(dx).$$

From this, it is easily deduced that the sequence  $(P_n)$  converges weakly to the random element  $(1 - \beta)\delta_h + \beta\delta_0$ , where  $\beta$  is a random variable with law  $\mathcal{L}$ , the common law of all the  $\beta_n$ . So we assume from now on that  $Q \neq \delta_0$ .

### 3. Main results

#### 3.1. Weak convergence

Recall that the sequence  $(P_n)$  in the random model has parameters  $(\beta_n)$ ,  $Q$ ,  $P_0$ , and  $h$ , with  $h = S_{P_0}$ . Then  $(P_n)$  converges weakly to a globally stable equilibrium, in the sense that the limit depends on  $P_0$  only through  $h$ . Recall that  $\beta$  is a random variable with law  $\mathcal{L}$ , the common law of the  $\beta_n$ .

**Theorem 2.** *For the random model (5), the sequence  $(P_n)$  converges weakly to a random probability measure, denoted by  $\mathcal{I}$ , whose distribution depends on  $\mathcal{L}$ ,  $Q$ ,  $h$  but not on the choice of  $P_0$ .*

**Remark 1.** In [23, p. 872], it is written that the distribution of  $\mathcal{I}$  depends on  $\beta$ ,  $Q$ ,  $h$ . This statement is true in the sense that the distribution of  $\mathcal{I}$  depends on  $\beta$  via its distribution. Here we make it clearer by replacing  $\beta$  by  $\mathcal{L}$ .

**Remark 2.** If we start with  $P_0 = \delta_h$  (recall that  $h \in [S_Q, 1]$ ), then all the  $P_n$  are supported on  $[0, S_Q] \cup \{h\}$ , which implies that the limit  $\mathcal{I}$  is supported on the same set  $[0, S_Q] \cup \{h\}$ . Moreover, we have either  $\mathcal{I}(h) > 0$  almost surely (a.s.) or  $\mathcal{I}(h) = 0$  a.s. (a justification is provided in Remark 12 in Section 4.4). In the latter case,  $\mathcal{I}$  is supported only on  $[0, S_Q]$ , and (the distribution of)  $\mathcal{I}$  does not depend on  $h$  (see Theorem 3). Therefore, although we say  $h$  is a parameter of  $\mathcal{I}$ , this should be understood in the sense that  $\mathcal{I}$  is the weak limit of  $(P_n)$  with  $h = S_{P_0}$ . The limit  $\mathcal{I}$  is introduced in Section 4.3, but the proof of weak convergence is deferred to a later stage, as it uses other main results, such as the condensation criterion for the random model.

**3.2. Condensation criterion**

The fact that either  $\mathcal{I}(h) > 0$  a.s. or  $\mathcal{I}(h) = 0$  a.s. allows us to give the precise definition of condensation, in line with that for Kingman’s model, as follows.

**Definition 1.** For the random model, we say there is condensation at the largest fitness value  $h$  if  $Q$  assigns zero mass at  $h$  (i.e.,  $Q(h) = 0$ ) but the limiting measure  $\mathcal{I}$  assigns positive mass at  $h$  (i.e.,  $\mathcal{I}(h) > 0$ , a.s.).

Next we give the condensation criterion. If  $h = S_Q$ , we write  $\mathcal{I}_Q$  for  $\mathcal{I}$  and  $\mathcal{K}_Q$  for  $\mathcal{K}$ .

**Theorem 3.** (Condensation criterion.) If there is no condensation at  $h$ , then  $\mathcal{I} \stackrel{d}{=} \mathcal{I}_Q$ . The condensation criterion for  $\mathcal{I}$  at  $h$  is as follows:

- 1. If  $h = S_Q$ , then there is no condensation at  $h$  if

$$\mathbb{E} \left[ \ln \frac{S_Q(1 - \beta)}{\int y \mathcal{I}_Q(dy)} \right] < 0. \tag{6}$$

- 2. If  $h > S_Q$ , then there is no condensation at  $h$  if and only if

$$\mathbb{E} \left[ \ln \frac{h(1 - \beta)}{\int y \mathcal{I}_Q(dy)} \right] \leq 0. \tag{7}$$

Here

$$\mathbb{E} \left[ \ln \frac{1 - \beta}{\int y \mathcal{I}_Q(dy)} \right]$$

is well defined, takes values in  $[-\infty, -\ln \int y Q(dy)]$ , and depends only on the marginal distributions of  $\beta$  and  $\mathcal{I}_Q$ .

**Remark 3.** In fact, if there is no condensation at  $h$ , then  $\mathcal{I}, \mathcal{I}_Q$  are the same random probability measure, based on the definition of  $\mathcal{I}$  introduced at the end of Section 4.3. But since here we do not have the definition yet, we write the weaker statement  $\mathcal{I} \stackrel{d}{=} \mathcal{I}_Q$ .

**Remark 4.** As the distribution of  $\mathcal{I}_Q$  is determined by  $Q$  and  $\mathcal{L}$  (the distribution of  $\beta$ ),

$$\mathbb{E} \left[ \ln \frac{1 - \beta}{\int y \mathcal{I}_Q(dy)} \right]$$

depends only on  $Q$  and  $\mathcal{L}$ . By Remark 13 in Section 4.4, we can only have

$$\mathbb{E} \left[ \ln \frac{S_Q(1 - \beta)}{\int y \mathcal{I}_Q(dy)} \right] \leq 0.$$

For the occurrence of condensation in the case where  $h = S_Q$ , the fact that we cannot say anything when

$$\mathbb{E} \left[ \ln \frac{S_Q(1 - \beta)}{\int y \mathcal{I}_Q(dy)} \right] = 0$$

can be better understood in Kingman’s model, which is a special random model. In this model,

$$\mathbb{E} \left[ \ln \frac{S_Q(1 - \beta)}{\int y \mathcal{I}_Q(dy)} \right] = 0$$

becomes

$$\ln \frac{S_Q(1-b)}{\int y\mathcal{K}_Q(dy)} = 0.$$

By some simple computations using Theorem 1, the above display is equivalent to

$$\int \frac{Q(dx)}{1-x/S_Q} \leq b^{-1}.$$

But it covers cases with and without condensation. For full details see Appendix A, where the case  $h > S_Q$  is also analysed.

We give some intuition for why Theorem 3 holds. Consider the unnormalised variant of the dynamical system that is given by

$$\bar{P}_n(dx) = (1 - \beta_n)x\bar{P}_{n-1}(dx) + \beta_n \left( \int y\bar{P}_{n-1}(dy) \right) Q(dx) \quad (8)$$

with  $\bar{P}_0 = P_0$ . By induction, it can be shown that

$$\bar{P}_n = P_n \prod_{i=0}^{n-1} \int xP_i(dx) \quad \forall n \geq 0. \quad (9)$$

We can roughly think of the growth of  $\bar{P}_n$  as contributed by two parts, the initial  $P_0$  and the subsequently arriving distributions  $Q$ . If the initial distribution is supported on  $[0, S_Q]$ , by Theorem 2,  $P_i$  converges weakly to  $\mathcal{I}_Q$  as  $i \rightarrow \infty$ . Then the part of  $\bar{P}_n$  contributed by the  $Q$ s grows at rate  $\text{gr}(Q) := \mathbb{E}[\ln \int x\mathcal{I}_Q]$  (see (9)). In comparison, the largest fitness value  $h$  in  $P_0$  can be assigned the growth rate  $\text{gr}(h) := \mathbb{E}[\ln h(1 - \beta)]$  (due to the term  $(1 - \beta_n)x\bar{P}_{n-1}(dx)$  in (8)). Then it is clear that the occurrence of condensation is determined by the comparison of  $\text{gr}(h)$  and  $\text{gr}(Q)$ . However, it is subtle when  $\text{gr}(h) = \text{gr}(Q)$ : there is no condensation if  $h > S_Q$ , and it is undetermined if  $h = S_Q$ .

In the follow-up paper [23], we provide a matrix representation for  $\mathcal{I}_Q$ , so the condensation criterion can be written neatly (see [23, Corollary 2, p. 877]). Moreover, using matrix analysis, we can compare the fitness of equilibria from different models (see [23, Section 3.3-(3), pp. 878–879]). The challenging problem of finding a necessary and sufficient condition for the occurrence of condensation in the case  $h = S_Q$  has not been dealt with anywhere and still remains open.

### 3.3. Invariant measure

We introduce the notion of *invariant measure*, which includes the limit  $\mathcal{I}$ . We will use invariant measures heavily in the proofs.

**Definition 2.** (Invariant measure.) *A random probability measure  $\nu$  is invariant if it is supported on  $[0, 1]$  and satisfies*

$$\nu(dx) \stackrel{d}{=} (1 - \beta) \frac{x\nu(dx)}{\int y\nu(dy)} + \beta Q(dx), \quad (10)$$

where  $\beta$  is independent of  $\nu$ . Clearly  $\mathcal{I}$  is an invariant measure, since it is the weak limit of  $(P_n)$  defined by (5).

**Theorem 4.** (Compoundness of invariant measures.) *For any invariant measure  $\nu$ , there exists a regular conditional distribution of  $\nu$  on  $S_\nu$ . Moreover, conditional on  $S_\nu$ ,*

$$(\nu|S_\nu) \stackrel{d}{=} \mathcal{I}, \quad \text{a.s.},$$

where  $\mathcal{I}$  is the random probability measure introduced in Theorem 2 with parameters  $\mathcal{L}, Q, h = S_\nu$  and satisfies  $\mathbb{P}(S_{\mathcal{I}} = S_\nu|S_\nu) = 1$ , a.s.

**Remark 5.** Remark 2 says that if there is no condensation at  $h$ , then  $\mathcal{I}$  is supported on  $[0, S_Q]$ . Since  $\mathcal{I}$  is an invariant measure, the above theorem implies that  $\mathcal{I} \stackrel{d}{=} \mathcal{I}_Q$ . This assertion has been stated in Theorem 3.

Using the notion of invariant measures, we can solve a distributional equation in the following example. For a survey on distributional equations, we refer to Aldous and Bandyopadhyay [1].

**Example 1.** Consider a particular case:  $Q$  is supported only on  $\{c\}$  for some  $c \in (0, 1)$ , and  $h \in (c, 1)$ . Let  $\nu$  be an invariant measure supported on  $\{c\} \cup \{h\}$ . Then  $\nu$  can be written as  $\nu = X\delta_c + (1 - X)\delta_h$ , where  $X$  is a random variable taking values in  $[0, 1]$  and satisfies

$$X\delta_c + (1 - X)\delta_h \stackrel{d}{=} (1 - \beta) \frac{cX\delta_c + h(1 - X)\delta_h}{cX + h(1 - X)} + \beta\delta_c,$$

with  $\beta$  independent of  $X$ . The above display is equivalent to

$$X \stackrel{d}{=} \frac{c + (h\beta - c)(1 - X)}{c + (h - c)(1 - X)}.$$

We are interested in a necessary and sufficient condition for the above equation to have a solution  $X$  with  $0 \leq X < 1$  a.s. (i.e.,  $\nu(h) > 0$  a.s.). By Theorem 4, this is equivalent to saying that there is condensation at  $h$ . By Theorem 3, the necessary and sufficient condition is simply  $\mathbb{E}[\ln(h(1 - \beta)/c)] > 0$ . Moreover, as such  $\nu$  is unique (in distribution), the solution  $X$  is also unique (in distribution).

The rest of the paper is organised as follows. Sections 4.1 and 4.2 provide necessary preparations. Sections 4.3 and 4.4 analyse the finite backward sequence, which is the main tool used in this paper. Section 4.5 proves Theorem 3. Section 4.4 analyses the invariant measures, and the results obtained there will be used in Section 4.7 to prove the weak convergence in Theorem 2. Section 4.8 is dedicated to the proof of Theorem 4.

## 4. Proofs

### 4.1. Relations between measures

We introduce the following notation to describe relations between measures:

1. For measures  $u, \nu \in M$ , we say  $u$  is a component of  $\nu$  on  $[0, a]$  (resp.  $[0, a)$ ), and we write  $u \leq_a \nu$  (resp.  $\leq_{a-}$ ), if

$$u(A) \leq \nu(A) \quad \text{for any measurable set } A \subset [0, a] \text{ (resp. } [0, a)).$$

For random measures  $\mu, \nu \in M$ , we write  $\mu \leq_a^d \nu$  if there exists a coupling  $(\mu', \nu')$  with  $\mu', \nu' \in M$  such that

$$\mu' \leq_a \nu' \text{ a.s. and } \mu' \stackrel{d}{=} \mu, \nu' \stackrel{d}{=} \nu. \quad (11)$$

The relation  $\mu \leq_{a-}^d \nu$  is defined in a similar way.



2. For measures  $(u_n)$  and  $u$  in  $M$ , we introduce the notation

$$u_n \leq_a \xrightarrow{TV} u,$$

which means that  $u_n \leq_a u_{n+1}$  for any  $n$ , and  $u_n$  converges in total variation to  $u$ . We define similarly  $\leq_{a-} \xrightarrow{TV}$ .

3. For real-valued random variables  $\xi, \eta$ , we write the well-known *stochastic ordering* as  $\xi \preceq \eta$ , which holds if

$$\mathbb{P}(\xi \leq x) \geq \mathbb{P}(\eta \leq x) \quad \forall x \in \mathbb{R}.$$

4. For any  $u \in M_1$ , let the distribution function of  $u$  be

$$D_u(x) := u([0, x]) \quad \forall x \in [0, 1].$$

For any  $u, v \in M_1$ , we use the same notation  $\preceq$  for the stochastic ordering and write  $u \preceq v$  if  $D_u(x) \geq D_v(x)$  for any  $x \in [0, 1]$ . This definition is natural, as  $\xi \preceq \eta$  is equivalent to  $u \preceq v$ , if  $u$  is the distribution of  $\xi$  and  $v$  is the distribution of  $\eta$ .

**Remark 6.** We make a comment about the relationship between  $\leq_{a-}$  and  $\preceq$ . For two probability measures  $u, v \in M_1$ , assume that  $S_u = S_v = a$ ; then  $u \leq_{a-} v$  implies that  $v \preceq u$ . But the converse is not true.

**Remark 7.** If we use the notation  $\leq_a, \leq_{a-}, \leq_a \xrightarrow{TV}, \leq_{a-} \xrightarrow{TV}, \preceq$  to describe the relations between random measures, it should be understood that they hold in the almost sure sense, or even in the pointwise sense (i.e., for every  $\omega \in \Omega$ ), if possible.

Similarly, if we use  $\leq, <, \geq, >, =, \neq$  to compare random variables, they should be understood in the almost sure sense or in the pointwise sense.

### 4.2. Three sequences

To study the asymptotic behaviour of  $(P_n)$ , we also introduce  $(P'_n), (P''_n)$  so that the three forward sequences correspond respectively to

$$((\beta_n), Q, P_0, h), \quad ((\beta_n), Q', P'_0, h'), \quad ((\beta_n), Q'', P''_0, h'').$$

The parameters of  $(P'_n)$  and  $(P''_n)$  will be specified when they are used. The two sequences will converge weakly when they are used, and  $(P_n)$  will be compared to them (or to one of them) to show that  $(P_n)$  also converges weakly. The first place where this technique is used is in Section 4.4.

Using (2), we write

$$P_n(dx) = \mathcal{M}_n(dx) + \mathcal{W}_n(dx) \tag{12}$$

with

$$\mathcal{M}_n(dx) = \left( \prod_{l=0}^{n-1} \frac{1 - \beta_{l+1}}{\int y P_l(dy)} \right) x^n P_0(dx)$$

and

$$\mathcal{W}_n(dx) = \sum_{j=1}^n \left( \prod_{l=j}^{n-1} \frac{1 - \beta_{l+1}}{\int y P_l(dy)} \right) b_j m_{n-j} Q^{n-j}(dx).$$

Therefore  $\mathcal{M}_n$  is the contribution to  $P_n$  made by  $P_0$ , while  $\mathcal{W}_n$  is the contribution by the  $Q$ s.

Similarly we introduce

$$P'_n(dx) = \mathcal{M}'_n(dx) + \mathcal{W}'_n(dx), \tag{13}$$

$$P''_n(dx) = \mathcal{M}''_n(dx) + \mathcal{W}''_n(dx) \tag{14}$$

with  $\mathcal{M}'_n, \mathcal{W}'_n, \mathcal{M}''_n, \mathcal{W}''_n$  defined correspondingly.

### 4.3. Introducing the finite backward sequences

4.3.1. *The general model.* We introduce the finite backward sequence  $(P^n_j) = (P^n_j)_{0 \leq j \leq n}$  for the general model which has parameters  $n, (b_j)_{1 \leq j \leq n}, Q, P^n_n, h$  with  $S_{P^n_n} = h$ :

$$P^n_j(dx) = (1 - b_{j+1}) \frac{xP^n_{j+1}(dx)}{\int yP^n_{j+1}(dy)} + b_{j+1}Q(dx) \quad \forall 0 \leq j \leq n - 1. \tag{15}$$

Here  $h, Q$  are from the general model and  $P^n_n$  can be any measure in  $M_1$  satisfying  $S_{P^n_n} = h$ . The  $(b_j)_{1 \leq j \leq n}$  are the first  $n$  mutation probabilities in the general model. Here we use the index  $j$  to indicate that we are dealing with a finite backward sequence.

The sequence is *backward* in the sense that we use  $b_n$  to generate  $P^n_{n-1}$  from  $P^n_n$ , use  $b_{n-1}$  to generate  $P^n_{n-2}$  from  $P^n_{n-1}$ , etc. That is, the  $(b_j)$  are used backwards and the  $(P^n_j)$  are generated backwards. The advantage in taking the backward approach is that  $(P^n_j)$  converges as  $n$  tends to infinity, in contrast to the forward sequence.

**Lemma 1.** *In the general model, for the finite backward sequence with  $P^n_n = \delta_h, P^n_j$  converges in total variation to a limit, denoted by  $\mathcal{G}_j = \mathcal{G}_{j,h}$ , as  $n$  goes to infinity with  $j$  fixed, such that*

$$\mathcal{G}_{j-1}(dx) = (1 - b_j) \frac{x\mathcal{G}_j(dx)}{\int y\mathcal{G}_j(dy)} + b_jQ(dx), \quad j \geq 1. \tag{16}$$

As a consequence,  $\mathcal{G}_0: [0, 1)^\infty \rightarrow M_1$  is a measurable function, with  $\mathcal{G}_j = \mathcal{G}_0(b_{j+1}, b_{j+2}, \dots)$  supported on  $[0, S_Q] \cup \{h\}$  for any  $j \geq 0$ .

**Remark 8.** We write  $\mathcal{G}_{j,h}$  when  $h$  has to be specified for clarity. Otherwise we write  $\mathcal{G}_j$ . This kind of abbreviation applies to other terms which will appear later.

**Remark 9.** Note that, by (16), the  $\mathcal{G}_j(h)$  are either all zero or all strictly positive.

*Proof.* We prove a stronger version below:

$$\text{For any } j, \quad P^n_j \xrightarrow{\text{TV}}_{h-} \mathcal{G}_j, \text{ as } n \rightarrow \infty. \tag{17}$$

It suffices to show that

$$P^n_j \leq_{h-} P^{n+1}_j, \tag{18}$$

as the  $P^n_j$  are all supported on  $[0, S_Q] \cup \{h\}$ .

First of all,  $P^n_n = \delta_h \leq_{h-} P^{n+1}_n$ . Assume that for some  $1 \leq j \leq n$  we have  $P^n_j \leq_{h-} P^{n+1}_j$ . By definition

$$P^n_{j-1}(dx) = (1 - b_j) \frac{xP^n_j(dx)}{\int yP^n_j(dy)} + b_jQ(dx), \quad P^{n+1}_{j-1}(dx) = (1 - b_j) \frac{xP^{n+1}_j(dx)}{\int yP^{n+1}_j(dy)} + b_jQ(dx). \tag{19}$$

Since  $P^n_j \leq_{h-} P^{n+1}_j$  (and hence  $P^{n+1}_j \leq P^n_j$ ; see Remark 6), we have

$$\int yP^{n+1}_j(dy) \leq \int yP^n_j(dy),$$

and thus

$$\frac{x}{\int yP_j^n(dy)} \leq \frac{x}{\int yP_j^{n+1}(dy)} \quad \forall x \in [0, 1].$$

From  $P_j^n \leq_{h-} P_j^{n+1}$  and (19), we get  $P_{j-1}^n \leq_{h-} P_{j-1}^{n+1}$ . Induction shows that

$$P_j^n \leq_{h-} P_j^{n+1} \quad \text{for any } 0 \leq j \leq n, \quad n \geq 0. \tag{20}$$

This completes the proof. □

The monotonicity analysis in the above proof will be used many times in this paper, as it applies to both backward and forward sequences. An immediate application is the following: we can compare  $(\mathcal{G}_j)$  and  $(\mathcal{G}'_j) = (\mathcal{G}_{j,h'})$  for  $S_Q \leq h < h' \leq 1$  with the same  $(b_j)$ ,  $Q$ .

**Corollary 1.** *Let  $(\mathcal{G}_j)$  and  $(\mathcal{G}'_j)$  be the above sequences. Then we have*

$$\mathcal{G}'_j \leq_{h-} \mathcal{G}_j, \quad \mathcal{G}_j(h) \leq \mathcal{G}'_j(h'), \quad \forall j \geq 0. \tag{21}$$

Moreover, we have the exact equalities in the above display for any  $h \in [S_Q, h']$  if and only if  $\mathcal{G}'_0(h') = 0$ . In this case  $(\mathcal{G}_j)$  and  $(\mathcal{G}'_j)$  are all supported on  $[0, S_Q]$ , and are both equal to  $(\mathcal{G}_{j,S_Q})$ .

*Proof.* Let  $(P_j^n)$  be the sequence in Lemma 1. Let  $(P_{j,h'}^n)$  be the variant of  $(P_j^n)$  with  $P_n^n = \delta_{h'}$ . By following the same monotonicity analysis as in the proof of (18), we obtain

$$P_{j,h'}^n \leq_{h-} P_j^n, \quad P_j^n(h) \leq P_{j,h'}^n(h'), \quad \forall 0 \leq j \leq n.$$

By Lemma 1,  $P_{j,h'}^n \xrightarrow{TV} \mathcal{G}'_j$  and  $P_j^n \xrightarrow{TV} \mathcal{G}_j$  as  $n \rightarrow \infty$ . Thus we obtain (21).

Now let us prove the if-and-only-if statement. If  $\mathcal{G}'_0(h') = 0$ , then by (21),  $\mathcal{G}_0(h) = 0$ . Using Remark 9,  $\mathcal{G}'_j(h') = 0$ ,  $\mathcal{G}_j(h) = 0$  for any  $j$ , and so (21) holds with equalities. For the other direction, if  $\mathcal{G}'_0(h') > 0$ , then again by Remark 9,  $\mathcal{G}'_j(h') > 0$  for any  $j$ . Using (21), it holds that

$$\int y\mathcal{G}'_j(dy) > \int y\mathcal{G}_j(dy) \quad \forall j. \tag{22}$$

Similarly to (19),

$$\mathcal{G}_{j-1}(dx) = (1 - b_j) \frac{x\mathcal{G}_j(dx)}{\int y\mathcal{G}_j(dy)} + b_jQ(dx), \quad \mathcal{G}'_{j-1}(dx) = (1 - b_j) \frac{x\mathcal{G}'_j(dx)}{\int y\mathcal{G}'_j(dy)} + b_jQ(dx).$$

As (22) implies that

$$\frac{1 - b_j}{\int y\mathcal{G}_j(dy)} > \frac{1 - b_j}{\int y\mathcal{G}'_j(dy)},$$

and again using (21), we obtain  $\mathcal{G}'_{j-1} \leq_{h-} \mathcal{G}_{j-1}$ , but they are not equal on  $[0, h)$ . Since they are probability measures, we have  $\mathcal{G}_{j-1}(h) < \mathcal{G}'_{j-1}(h')$  for any  $j$ . This completes the proof.

If (21) holds with equalities,  $(\mathcal{G}_j) = (\mathcal{G}'_j)$  are all supported on  $[0, S_Q]$ . To show that they are equal to  $(\mathcal{G}_{j,S_Q})$ , we only have to take  $h = S_Q$  and apply the equalities in (21). □

4.3.2. *The random model.* The goal of this paper is the random model, which is a randomised general model. Since  $(\mathcal{G}_j)$  has parameters  $(b_{j+1}, b_{j+2}, \dots)$  and  $Q, h$ , we can define

$$\mathcal{I}_j = \mathcal{I}_{j,h} := \mathcal{G}_0(\beta_{j+1}, \beta_{j+2}, \dots).$$

Therefore  $\mathcal{I}_j$  is the quenched limit of the finite backward sequences in the random model with  $P^n = \delta_h$ . Thanks to Lemma 1, we have the following result.

**Corollary 2.** *The sequence  $(\mathcal{I}_j) = (\mathcal{I}_j)_{j \geq 0}$  is stationary ergodic and satisfies*

$$\mathcal{I}_{j-1}(dx) = (1 - \beta_j) \frac{x\mathcal{I}_j(dx)}{\int y\mathcal{I}_j(dy)} + \beta_j Q(dx), \quad j \geq 1. \tag{23}$$

**Remark 10.** The equality (23) holds in the pointwise sense. In other words, given any realisation of  $(\beta_j)$  (or equivalently, conditioning on  $(\beta_j)$ ), the equality holds for any  $j$  as in the general model. In the sequel, when we present results regarding  $(\mathcal{I}_j)$ , then *conditioning on  $(\beta_j)$*  should be understood as *in the pointwise sense*. Sometimes we omit to state either of the two phrases when the context is clear.

The proof of Corollary 2 requires the following result, which is proved in [15, Lemma 9.5].

**Lemma 2.** *Let  $(S, \mathcal{S})$  and  $(S', \mathcal{S}')$  be measurable spaces. Let  $(\alpha_j) \in S^\infty$  be a stationary ergodic sequence of random variables. Let  $f: S^\infty \rightarrow S'$  be a measurable function. Then  $(f(\alpha_j, \alpha_{j+1}, \dots))$  is also stationary ergodic.*

*Proof of Corollary 2.* Since  $(\beta_j)$  is i.i.d., it is stationary ergodic. As  $\mathcal{G}_0$  is a measurable function from  $[0, 1)^\infty$  to  $M_1$ , we apply Lemma 2 to obtain that  $(\mathcal{I}_j) = (\mathcal{G}_0(\beta_{j+1}, \beta_{j+2}, \dots))$  is also stationary ergodic. The recursive equation (23) is inherited from (16). □

Since  $(\mathcal{I}_j)$  is stationary ergodic, all of the  $\mathcal{I}_j$  have the same distribution. We define  $\mathcal{I} := \mathcal{I}_0 = \mathcal{I}_{0,h}$ , which is the weak limit appearing in Theorem 2. The reason we drop the index is to distinguish this from the backward context, when it is appropriate to do so. The term  $\mathcal{I}_Q$  used in Theorem 3 is in fact  $\mathcal{I}_{0,S_Q}$ .

We comment further on the importance of finite backward sequences. Let  $(P_n)$  be a forward sequence and  $(P^n_j)$  the finite backward sequence with  $P^n_n = P_0$ , both in the random model with the same  $(\beta_j)$  and  $Q$ . Since  $(\beta_j)$  is i.i.d., we have

$$(P_0, P_1, \dots, P_n) \stackrel{d}{=} (P^n_n, P^n_{n-1}, \dots, P^n_0). \tag{24}$$

So showing the weak convergence of  $(P_n)_{n \geq 0}$  is equivalent to showing that of  $(P^n_0)_{n \geq 0}$ . But investigating the finite backward sequences, via the general model, appears to be more convenient. In general a dynamical system is easier to handle if we take a backward point of view; see Diaconis and Freedman [9].

### 4.4. Finer analysis of the finite backward sequences

4.4.1. *The general model.* We consider  $(P^n_j)$  with  $P^n_n = \delta_h$ , as in Lemma 1. Developing (15), we obtain

$$P^n_0(dx) = \left( \prod_{l=1}^n \frac{1 - b_l}{\int y P^n_l(dy)} \right) x^n P^n_n(dx) + \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int y P^n_l(dy)} \right) b_{j+1} m_j Q^j(dx) \tag{25}$$

$$= \left( \prod_{l=1}^n \frac{h(1 - b_l)}{\int y P^n_l(dy)} \right) \delta_h(dx) + \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int y P^n_l(dy)} \right) b_{j+1} m_j Q^j(dx). \tag{26}$$

We refer to (2) for the expansion of the forward sequence  $(P_n)$ .

**Proposition 1.** Let  $(P_j^n)$  be the finite backward sequence in the general model with  $P_n^n = \delta_h$ . Then for the sequence  $(G_j)$ , we have

$$G_0(dx) = G_0\delta_h(dx) + \sum_{j=0}^{\infty} \prod_{l=1}^j \frac{(1 - b_l)}{\int yG_l(dy)} b_{j+1}m_jQ^j(dx), \tag{27}$$

where the second term on the right side of (26) converges to that of (27),

$$\sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int yP_l^n(dy)} \right) b_{j+1}m_jQ^j(dx) \leq_{S_Q^-} \xrightarrow{TV} \sum_{j=0}^{\infty} \prod_{l=1}^j \frac{(1 - b_l)}{\int yG_l(dy)} b_{j+1}m_jQ^j(dx), \tag{28}$$

and the term  $G_0 = G_{0,h}$  satisfies the following assertions:

$$\prod_{l=1}^n \frac{h(1 - b_l)}{\int yP_l^n(dy)} \text{ decreases in } n \text{ and converges to } G_0; \tag{29}$$

$$G_0 = 1 - \sum_{j=0}^{\infty} \prod_{l=1}^j \frac{(1 - b_l)}{\int yG_l(dy)} b_{j+1}m_j \in [0, 1]; \tag{30}$$

$$G_0 = G_0(h) \text{ if } Q(h) = 0; \tag{31}$$

$$\int \left(\frac{y}{h}\right)^n G_n(dy) \prod_{l=1}^n \frac{h(1 - b_l)}{\int yG_l(dy)} \text{ decreases in } n \text{ and converges to } G_0, \text{ if } G_0 > 0. \tag{32}$$

Moreover, if we define  $G_j$  for  $G_j$  similarly to  $G_0$  for  $G_0$ , we have

$$G_{j-1} = G_j \frac{h(1 - b_j)}{\int yG_j(dy)} \quad \forall j \geq 1. \tag{33}$$

As a consequence, the  $G_j$  are either all 0 or all strictly positive.

*Proof.* By (17),  $\int yP_l^n(dy)$  increases in  $n$  and converges to  $\int yG_l(dy)$  as  $n \rightarrow \infty$ . Then, using (26), we obtain (28). Integrating on both sides of (26), we use (28) to deduce that

$$\prod_{l=1}^n \frac{h(1 - b_l)}{\int yP_l^n(dy)} = 1 - \int \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int yP_l^n(dy)} \right) b_{j+1}m_jQ^j(dx)$$

decreases in  $n$  and converges to the limit

$$1 - \int \sum_{j=0}^{\infty} \left( \prod_{l=1}^j \frac{1 - b_l}{\int yG_l(dy)} \right) b_{j+1}m_jQ^j(dx) =: G_0.$$

So (27), (29), and (30) are proved.

From (27) we observe (31). To show (32), we develop (16) as follows:

$$\begin{aligned} G_0(dx) &= \left( \prod_{l=1}^n \frac{1 - b_l}{\int yG_l(dy)} \right) x^n G_n(dx) + \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int yG_l(dy)} \right) b_{j+1}m_jQ^j(dx) \\ &= \left( \int \left(\frac{y}{h}\right)^n G_n(dy) \prod_{l=1}^n \frac{h(1 - b_l)}{\int yG_l(dy)} \right) \frac{x^n G_n(dx)}{\int y^n G_n(dy)} + \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1 - b_l}{\int yG_l(dy)} \right) b_{j+1}m_jQ^j(dx). \end{aligned}$$

Combining the above display with (27) and (28), we obtain (32), and also that

$$\frac{x^n \mathcal{G}_n(dx)}{\int y^n \mathcal{G}_n(dy)}$$

converges weakly to  $\delta_h$ . Finally, combining (16) and (27) leads to (33). □

**Remark 11.** The proposition implies that  $(P_n^n)$  with  $P_n^n = \delta_h$  in the random model converges in total variation to  $\mathcal{I} = \mathcal{I}_0$ , pointwise. Then by (24),  $(P_n)$  in the random model with  $P_0 = \delta_h$  converges weakly to  $\mathcal{I}$ . Therefore Theorem 2 is proved for the particular case with  $P_0 = \delta_h$ . As will become clear later (in Section 4.7), a complete proof has to deal with different kinds of  $P_0$ . The proof here with  $P_0 = \delta_h$  is the simplest case.

4.4.2. *The random model.* When carrying over the results of Proposition 1 to the random model, we change the symbol  $G$  to  $I$ , analogously to the change from  $\mathcal{G}$  to  $\mathcal{I}$ . For instance, we set  $I_j = G_0(\beta_{j+1}, \beta_{j+2}, \dots)$  for any  $j \geq 0$ . Then we have the following corollary.

**Corollary 3.** *The process  $(I_j) = (I_j)_{j \geq 0}$  is stationary ergodic. Moreover,  $\mathbb{P}(\{I_j = 0, \forall j\}) = \mathbb{P}(I_0 = 0) \in \{0, 1\}$ .*

**Remark 12.** If  $Q(h) > 0$ , then it must be that  $h = S_Q$  and  $\mathcal{I}(h) = \mathcal{I}_0(h) > 0$  a.s. If  $Q(h) = 0$ , then  $\mathcal{I}(h) = \mathcal{I}_0(h) = I_0$ . So applying Corollary (3), either  $\mathcal{I}(h) > 0$  a.s., or  $\mathcal{I}(h) = 0$  a.s.

*Proof.* By Proposition 1,  $G_0 = G_0(b_1, b_2, \dots)$  is a measurable function from  $[0, 1)^\infty$  to  $[0, 1]$ . As  $(\beta_j)$  is i.i.d., we obtain that  $(I_j) = (G_0(\beta_{j+1}, \beta_{j+2}, \dots))$  is stationary ergodic, thanks to Lemma 2.

By (33), for any  $k$ ,  $\{I_k = 0\} = \{I_j = 0, \forall j\}$ . Note that  $\{I_j = 0, \forall j\}$  is an invariant set in the sigma-algebra generated by  $(I_j)$ . By ergodicity of  $(I_j)$ ,  $\mathbb{P}(\{I_j = 0, \forall j\}) = \mathbb{P}(I_0 = 0) \in \{0, 1\}$ . □

The following result provides a tool for finding out more about  $\mathcal{I}$  and  $Q$ . Let  $I = I_{0,h}$  and  $I_Q = I_{0,S_Q}$ . To summarise,  $\mathcal{I}, I, \mathcal{I}_Q, I_Q$  are identical in value to  $\mathcal{I}_{0,h}, I_{0,h}, \mathcal{I}_{0,S_Q}, I_{0,S_Q}$ , respectively.

**Corollary 4.** *The following statements about  $\mathbb{E} \left[ \ln \frac{1-\beta}{\int y \mathcal{I}(dy)} \right]$  hold:*

1.  $\mathbb{E} \left[ \ln \frac{1-\beta}{\int y \mathcal{I}(dy)} \right]$  is well defined, takes values in  $[-\infty, -\ln \int y Q(dy)]$ , and depends only on the marginal distributions of  $\beta$  and  $\mathcal{I}$ .

2. If  $Q(h) = 0$ , then

$$\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y \mathcal{I}(dy)} \right] \leq 0.$$

3. If  $\mathcal{I}(h) > 0$  a.s. and  $Q(h) = 0$ , then

$$\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y \mathcal{I}(dy)} \right] = 0.$$

4. If  $h = S_Q$  and  $Q(S_Q) > 0$ , then

$$\mathbb{E} \left[ \ln \frac{S_Q(1-\beta)}{\int y \mathcal{I}(dy)} \right] < 0 \text{ and } I = 0, \quad \text{a.s.}$$

**Remark 13.** If  $h = S_Q$ , we can only have

$$\mathbb{E} \left[ \frac{S_Q(1-\beta)}{\int y \mathcal{I}(dy)} \right] = \mathbb{E} \left[ \frac{S_Q(1-\beta)}{\int y \mathcal{I}_Q(dy)} \right] \leq 0.$$

*Proof.* Part 1: By (27),  $\mathcal{G}_0$  is a convex combination of probability measures  $\{\delta_h, Q, Q^1, Q^2, \dots\}$ . As  $Q^j \leq Q^{j+1} \leq \delta_h$  for any  $j \geq 0$ , we have, in the pointwise sense,

$$Q \leq \mathcal{I} = \mathcal{I}_0 \leq \delta_h. \tag{34}$$

Then

$$\ln \int yQ(dy) \leq \mathbb{E} \left[ \ln \int y\mathcal{I}(dy) \right] \leq \ln h.$$

So  $\mathbb{E} \left[ \ln \int y\mathcal{I}(dy) \right]$  is a finite term. Consequently,

$$\begin{aligned} \mathbb{E} \left[ \ln \frac{1 - \beta}{\int y\mathcal{I}(dy)} \right] &= \mathbb{E} \left[ \ln(1 - \beta) - \ln \int y\mathcal{I}(dy) \right] \\ &= \mathbb{E}[\ln(1 - \beta)] - \mathbb{E} \left[ \ln \int y\mathcal{I}(dy) \right] \in \left[ -\infty, -\ln \int yQ(dy) \right]. \end{aligned}$$

We observe that the above display depends only on the marginal distributions of  $\beta$  and  $\mathcal{I}$ .

Part 2: Let  $(P_j^n)$  be the finite backward sequence in the random model with  $P_n^n = \delta_h$ . By assumption,  $Q(h) = 0$ . Adapting (26) to the random model and taking the expectation of the mass on  $h$ , we obtain

$$1 \geq \mathbb{E}[P_0^n(h)] = \mathbb{E} \left[ \left( \prod_{l=1}^n \frac{h(1 - \beta_l)}{\int yP_l^n(dy)} \right) \right] \geq \exp \left( \sum_{l=1}^n \mathbb{E} \left[ \ln \frac{h(1 - \beta_l)}{\int yP_l^n(dy)} \right] \right),$$

where the second inequality is due to Jensen's inequality. By (17),

$$\mathbb{E} \left[ \ln \frac{h(1 - \beta_l)}{\int yP_l^n(dy)} \right] \text{ increases in } n \text{ and converges to } \mathbb{E} \left[ \ln \frac{h(1 - \beta_l)}{\int y\mathcal{I}_l(dy)} \right] = \mathbb{E} \left[ \ln \frac{h(1 - \beta)}{\int y\mathcal{I}(dy)} \right].$$

Combining the above two displays, it must be that

$$\mathbb{E} \left[ \ln \frac{h(1 - \beta)}{\int y\mathcal{I}(dy)} \right] \leq 0.$$

Part 3: Lemma 1 implies that there exists a measurable function  $T: [0, 1)^\infty \mapsto (0, \infty)$  such that for any  $j$ ,

$$\frac{h(1 - b_j)}{\int y\mathcal{G}_j(dy)} = T(b_j, b_{j+1}, \dots).$$

By Lemma 2,

$$\left( \frac{h(1 - \beta_j)}{\int y\mathcal{I}_j(dy)} \right) \text{ is stationary ergodic.}$$

By (32) and the fact that  $\mathcal{I}(h) = \mathcal{I}_0(h) = I_0 > 0$  a.s. (because  $Q(h) = 0$  by assumption), we have

$$\lim_{n \rightarrow \infty} (I_0)^{1/n} = \lim_{n \rightarrow \infty} \exp \left( \frac{1}{n} \ln \int \left( \frac{y}{h} \right)^n \mathcal{I}_n(dy) + \frac{1}{n} \sum_{l=1}^n \ln \frac{h(1 - \beta_l)}{\int y\mathcal{I}_l(dy)} \right) \stackrel{a.s.}{=} 1. \tag{35}$$

As  $(\mathcal{I}_j)$  is stationary ergodic,  $\int \left( \frac{y}{h} \right)^n \mathcal{I}_n(dy) \in [I_n, 1]$  converges weakly to  $I_0$ , which is strictly positive. Then

$$\left[ \frac{1}{n} \ln I_n, 0 \right] \ni \frac{1}{n} \ln \int \left( \frac{y}{h} \right)^n \mathcal{I}_n(dy) \xrightarrow{d} 0, \quad n \rightarrow \infty.$$

Moreover, since  $\left(\frac{h(1-\beta_l)}{\int y\mathcal{I}_l(dy)}\right)$  is stationary ergodic, we have

$$\frac{1}{n} \sum_{l=1}^n \ln \frac{h(1-\beta_l)}{\int y\mathcal{I}_l(dy)} \xrightarrow{a.s.} \mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}(dy)} \right], \quad n \rightarrow \infty.$$

The above three displays yield

$$1 = \exp \left( \mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}(dy)} \right] \right) \quad \text{or equivalently} \quad \mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}(dy)} \right] = 0.$$

Part 4: We show by contradiction that  $I = I_0 = 0$  a.s. Adapting (26) to the random model, we have

$$P_0^n(dx) = \left( \prod_{l=1}^n \frac{S_Q(1-\beta_l)}{\int yP_l^n(dy)} \right) \delta_{S_Q}(dx) + \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{1-\beta_l}{\int yP_l^n(dy)} \right) \beta_{j+1} m_j Q^j(dx).$$

If  $I_0 > 0$  a.s., we consider the mass on  $S_Q$  in the above display. Note that  $m_j Q^j(S_Q) = S_Q^j Q(S_Q)$ . By (29) we obtain

$$1 \geq P_0^n(S_Q) \geq Q(S_Q) \sum_{j=0}^{n-1} \left( \prod_{l=1}^j \frac{S_Q(1-\beta_l)}{\int yP_l^n(dy)} \right) \beta_{j+1} \geq Q(S_Q) \sum_{j=0}^{n-1} I_0 \beta_{j+1} \xrightarrow{d} \infty.$$

This is a contradiction. So  $I_0 = 0$  a.s. Note that by (34),  $\mathcal{I}(S_Q) = \mathcal{I}_0(S_Q) \geq Q(S_Q) > 0$ . Then we get  $\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}(dy)} \right] < 0$  using (35) and the arguments thereafter. □

### 4.5. Proof of Theorem 3

*Proof of Theorem 3.* The statement about  $\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}_Q(dy)} \right]$  concerns just a subcase of Corollary 4-1. So this is proved.

If there is no condensation at  $h$ , then by Corollary 1,  $\mathcal{I} = \mathcal{I}_{0,h} = \mathcal{I}_{0,S_Q} = \mathcal{I}_Q$ , so of course  $\mathcal{I} \stackrel{d}{=} \mathcal{I}_Q$ .

The first assertion in the condensation criterion holds by Corollary 4-3. We consider the second one. If there is condensation at  $h$ , then  $\mathcal{I}_{0,S_Q} \neq \mathcal{I}_{0,h}$ . By Corollary 1,  $\mathcal{I}_{0,h} \leq_{S_Q} \mathcal{I}_{0,S_Q}$  and  $\mathcal{I}_{0,S_Q}(S_Q) \leq \mathcal{I}_{0,h}(h)$ , which together with Corollary 4-3 implies that

$$\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}_{0,S_Q}(dy)} \right] > \mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}_{0,h}(dy)} \right] = 0.$$

The above inequality is strict because  $\mathcal{I}_{0,S_Q} \neq \mathcal{I}_{0,h}$ .

If there is no condensation at  $h$ , then by Corollary 1,  $\mathcal{I}_{0,h} = \mathcal{I}_{0,S_Q}$ . Using Corollary 4-2, we have

$$\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}_{0,S_Q}(dy)} \right] = \mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y\mathcal{I}_{0,h}(dy)} \right] \leq 0.$$



#### 4.6. Some properties of invariant measures

In this section, we prove some results concerning invariant measures. However, we leave the proof of Theorem 4 to the end. Invariant measures will play important roles in the proof of Theorem 2.

**Lemma 3.** For any invariant measure  $\nu$ ,

$$\mathbb{E} \left[ \ln \frac{1 - \beta}{\int y \nu(dy)} \right]$$

is well defined, takes values in  $[-\infty, -\ln \int y Q(dy)]$ , and depends only on the marginal distributions of  $\beta$  and  $\nu$ .

*Proof.* By the definition of invariant measure,

$$\begin{aligned} \mathbb{E} \left[ \int y \nu(dy) \right] &= (1 - \mathbb{E}[\beta]) \mathbb{E} \left[ \frac{\int y^2 \nu(dy)}{\int y \nu(dy)} \right] + \mathbb{E}[\beta] \int y Q(dy) \\ &\geq (1 - \mathbb{E}[\beta]) \mathbb{E} \left[ \int y \nu(dy) \right] + \mathbb{E}[\beta] \int y Q(dy), \end{aligned}$$

where the inequality is due to the fact that  $\int y^2 \nu(dy) \geq (\int y \nu(dy))^2$ . Then we obtain

$$\int y Q(dy) \leq \mathbb{E} \left[ \int y \nu(dy) \right] \leq 1.$$

Proceeding similarly as in the proof of Corollary 4–1, we conclude that this lemma holds.  $\square$

**Corollary 5.**  $\mathcal{I}_Q$  is the unique (in distribution) invariant measure supported on  $[0, S_Q]$ .

*Proof.* Let  $\nu$  be any invariant measure on  $[0, S_Q]$ . We show that  $\nu \stackrel{d}{=} \mathcal{I}_Q$ . Note that  $S_\nu = S_Q$ , a.s. Let  $(P_n)$  and  $(P'_n)$  be two forward sequences as in Section 4.2 with

$$Q = Q', \quad h = h' = S_Q, \quad P_0 \stackrel{d}{=} \nu, \quad P'_0 = \delta_{S_Q},$$

and with  $P_0$  independent of  $(\beta_n)$ . The two sequences differ only in the starting measures (satisfying  $P_0 \leq_{S_Q} P'_0$ ), with other parameters identical. Since  $\nu$  is invariant,  $P_n \stackrel{d}{=} \nu$  for any  $n \geq 0$ . Using the notation  $\mathcal{M}_n, \mathcal{M}'_n, \mathcal{W}_n, \mathcal{W}'_n$  from Section 4.2, and by a monotonicity analysis as in the proof of Lemma 1, we obtain, in the pointwise sense,

$$\int \mathcal{M}_n(dx) \leq \int \mathcal{M}'_n(dx), \quad \mathcal{W}'_n \leq_{S_Q} \mathcal{W}_n. \quad (36)$$

If  $I_Q = 0$  a.s., by (29) in Proposition 1 and (24),

$$\int \mathcal{M}'_n(dx) \xrightarrow{d} I_Q \stackrel{\text{a.s.}}{=} 0.$$

Remark 11 says that  $P'_n (= \mathcal{W}'_n + \mathcal{M}'_n) \xrightarrow{d} \mathcal{I}_Q$ . So

$$\mathcal{W}'_n \xrightarrow{d} \mathcal{I}_Q.$$

Thus, applying (36) and the fact that  $I_Q = 0$  a.s., we obtain

$$\int \mathcal{M}_n(dx) \xrightarrow{d} 0, \quad \mathcal{W}_n \xrightarrow{d} \mathcal{I}_Q.$$

Consequently,

$$P_n (= \mathcal{W}_n + \mathcal{M}_n) \xrightarrow{d} \mathcal{I}_Q.$$

Since  $\nu \stackrel{d}{=} P_n$  for any  $n$ , we have  $\nu \stackrel{d}{=} \mathcal{I}_Q$ .

If  $I_Q > 0$  a.s., then by Corollary 4-4,  $Q(S_Q) = 0$  and  $\mathcal{I}_Q(S_Q) = I_Q > 0$  a.s. Then by Corollary 4-3, we have

$$\mathbb{E} \left[ \ln \frac{S_Q(1-\beta)}{\int y \mathcal{I}_Q(dy)} \right] = 0. \quad (37)$$

Again using a monotonicity analysis, in a pointwise sense, we have

$$P'_n \leq_{S_Q^-} P_n, \quad P_n(S_Q) \leq P'_n(S_Q). \quad (38)$$

As  $P'_n \xrightarrow{d} \mathcal{I}_Q$ , and  $P_n \stackrel{d}{=} \nu$  for all  $n$ , the above display implies that

$$\mathcal{I}_Q \leq_{S_Q^-} \nu, \quad \nu(S_Q) \leq I_Q = \mathcal{I}_Q(S_Q).$$

Assume that  $\nu$  is not equal to  $\mathcal{I}_Q$  in distribution; then by the above display and (37),

$$\mathbb{E} \left[ \ln \frac{S_Q(1-\beta)}{\int y \nu(dy)} \right] > 0.$$

The inequality implies that for  $\varepsilon > 0$  small enough, we have

$$\mathbb{E} \left[ \ln \frac{(S_Q - \varepsilon)(1-\beta)}{\int y \nu(dy)} \right] > 0.$$

As  $S_\nu = S_Q$  a.s. and  $P_0 \stackrel{d}{=} \nu$ ,

$$\int \left( \frac{x}{S_Q - \varepsilon} \right)^n P_0(dx) \xrightarrow{d} \infty$$

as  $n \rightarrow \infty$ . Again using the decomposition (12) in Section 4.2, we get

$$\begin{aligned} 1 &= \mathbb{E} \left[ \int P_0(dx) \right] \geq \mathbb{E} \left[ \int \mathcal{M}_n(dx) \right] \\ &= \mathbb{E} \left[ \exp \left( \ln \int \left( \frac{x}{S_Q - \varepsilon} \right)^n P_0(dx) + \sum_{l=0}^{n-1} \ln \frac{(S_Q - \varepsilon)(1 - \beta_{l+1})}{\int y P_l^{\beta_l}(dy)} \right) \right] \\ &\geq \mathbb{E} \left[ \exp \left( \sum_{l=0}^{n-1} \ln \frac{(S_Q - \varepsilon)(1 - \beta_{l+1})}{\int y P_l^{\beta_l}(dy)} \right) \right] \geq \exp \left( n \mathbb{E} \left[ \frac{(S_Q - \varepsilon)(1 - \beta)}{\int y \nu(dy)} \right] \right) \xrightarrow{n \rightarrow \infty} \infty, \end{aligned}$$

where the third inequality is due to Jensen's inequality. So this is a contradiction, which means that  $\nu$  is equal in distribution to  $\mathcal{I}_Q$ .  $\square$

**4.7. Proof of Theorem 2**

**Case 1.**  $P_0 = \delta_h$ .

*Proof of Theorem 2, Case 1.* This is shown in Remark 11. □

**Case 2.**  $I_{0,h} = 0$  a.s.

*Proof of Theorem 2, Case 2.* Let  $(P_n)_{n \geq 0}, (P'_n)_{n \geq 0}$  be two forward sequences as in Section 4.2 with

$$Q = Q', \quad h = h', \quad P'_0 = \delta_h.$$

So the two sequences differ only in the starting measures (satisfying  $P_0 \leq_{h-} P'_0$ ), with other parameters identical. Next it suffices to follow the same procedure as in the proof of Corollary 5 for the case  $I_Q = 0$  a.s. The proof is omitted. □

**Case 3.**  $I_{0,h} > 0$  a.s. and  $P_0(h) > 0$ .

For Case 3, we first restate a result from [22, p. 10]. The statement in [22] considers only  $h = 1$ , but it is easily generalised to any  $h$ . Recall the distribution function  $D_u$  for  $u \in M_1$ , introduced in Section 4.1.

**Lemma 4.** *Let  $u_1, u_2 \in M_1$  be any probability measures satisfying  $S_{u_1} = S_{u_2} = h$  and  $u_1 \leq_{h-} u_2$ . If for some  $\varepsilon > 0$  there exists  $a \in (0, h)$  such that  $D_{u_1}(a) + \varepsilon \leq D_{u_2}(a)$ , then*

$$\int y u_1(dy) \geq c(a, \varepsilon) \int y u_2(dy),$$

where

$$c(a, \varepsilon) = \frac{1}{1 - \varepsilon(h - a)} > 1.$$

*Proof of Theorem 2, Case 3.* Let  $(P_n), (P'_n)$  be the two forward sequences in the proof of Case 2. Similarly to (38), conditionally on  $(\beta_n)$ , we have

$$P'_n \leq_{h-} P_n, \quad P_n(h) \leq P'_n(h), \tag{39}$$

implying

$$\int y P'_j(dy) \geq \int y P_j(dy) \quad \forall j \geq 0.$$

For any  $\varepsilon > 0, a \in (0, h)$ , let

$$\kappa_n := \#\{n: D_{P'_j}(a) + \varepsilon \leq D_{P_j}(a), 0 \leq j \leq n\}.$$

Note that by Proposition 1–4,  $Q(h) = 0$ . So using (12) and (13) in Section 4.2, we have

$$P'_n(h) = \prod_{l=0}^{n-1} \frac{h(1 - \beta_{l+1})}{\int y P'_l(dy)}, \quad P_n(h) = \left( \prod_{l=0}^{n-1} \frac{h(1 - \beta_{l+1})}{\int y P_l(dy)} \right) P_0(h).$$

Then by Lemma 4, we have

$$\prod_{l=0}^n \int y P'_l(dy) \geq c(a, \varepsilon)^{\kappa_n} \prod_{l=0}^n \int y P_l(dy).$$

Therefore,

$$P'_n(h) \leq \frac{1}{c(a, \varepsilon)^{\kappa_n}} \left( \prod_{l=0}^{n-1} \frac{h(1 - \beta_{l+1})}{\int y P_l(dy)} \right) = \frac{P_n(h)}{c(a, \varepsilon)^{\kappa_n} P_0(h)} \leq \frac{1}{c(a, \varepsilon)^{\kappa_n} P_0(h)}.$$

But (29) of Proposition 1 and (24) imply that  $P'_n(h)$  converges weakly to  $I_{0,h}$ , which is by assumption non-zero a.s. Thus  $\lim_{n \rightarrow \infty} \kappa_n < \infty$  a.s. As  $a, \varepsilon$  are arbitrary numbers and by Case 1 of this theorem  $P'_n \xrightarrow{d} \mathcal{I}_{0,h}$ , we use (39) to conclude that  $P_n$  also converges weakly to  $\mathcal{I}_{0,h}$ .  $\square$

**Case 4.**  $I_{0,h} > 0$  a.s. and  $P_0(h) = 0$ .

*Proof of Theorem 2, Case 4.* The idea is to use a tripling argument similarly as in the proof of Theorem 5 in [22]. For any  $u \in M_1$  and any  $a \in [0, 1]$ , define

$$u^a = u_{(0,a)} + u([a, 1])\delta_a, \quad a < h,$$

where  $u_{(0,a)}$  is the restriction of  $u$  on  $[0, a)$ .

We distinguish between  $h > S_Q$  and  $h = S_Q$ . For the former, let  $(P_n), (P'_n), (P''_n)$  be three forward sequences as in Section 4.2 with

$$S_Q < h'' < h = h'; \quad Q' = Q, \quad Q'' = Q^{h''} = Q; \quad P'_0 = \delta_h, \quad P''_0 = P_0^{h''}.$$

So the three sequences differ in the starting measures, including the largest fitness values, but have the same  $Q$  and  $(\beta_n)$ . Since  $P'_0(h') = \delta_h(h) = 1$  and  $0 < P''_0(h'') \leq 1$ , we use Case 1 for  $(P'_n)$  and Cases 2–3 for  $(P''_n)$  to obtain that

$$P'_n \xrightarrow{d} \mathcal{I}_{0,h}, \quad P''_n \xrightarrow{d} \mathcal{I}_{0,h''}. \tag{40}$$

Applying the monotonicity analysis, we find that the following holds in the pointwise sense:

$$P'_n \leq_{h-} P_n \leq_{h''-} P''_n. \tag{41}$$

Letting  $h'' \rightarrow h$  and using Corollary 1, we have that conditionally on  $(\beta_j)$ ,  $\mathcal{I}_{0,h''}$  converges weakly to a limit in  $M_1$ , denoted by  $\nu$ . So  $\nu$  is a (pointwise) weak limit of  $\mathcal{I}_{0,h''}$  as  $h'' \rightarrow h$ . We prove next that  $\nu = \mathcal{I}_{0,h}$ .

Since  $I_{0,h} > 0$  a.s. and  $h > S_Q$ , by Theorem 3,

$$\mathbb{E} \left[ \ln \frac{h(1 - \beta)}{\int y \mathcal{I}_{0,S_Q}(dy)} \right] > 0.$$

Then for  $h''$  close enough to  $h$ , we also have

$$\mathbb{E} \left[ \ln \frac{h''(1 - \beta)}{\int y \mathcal{I}_{0,S_Q}(dy)} \right] > 0.$$

The above display implies that there is condensation at  $h''$ , thanks to Theorem 3. Together with Corollary 4–3, this gives us

$$\mathbb{E} \left[ \ln \frac{h''(1 - \beta)}{\int y \mathcal{I}_{0,h''}(dy)} \right] = 0, \quad I_{0,h''} > 0, \quad \text{a.s.} \tag{42}$$

Since  $\mathcal{I}_{0,h''}$  is an invariant measure, the limit  $\nu$  is still an invariant measure. By (42) and Corollary 1, the pointwise convergence of  $\mathcal{I}_{0,h''}$  to  $\nu$  as  $h'' \rightarrow h$  implies

$$\mathbb{E} \left[ \ln \frac{h(1 - \beta)}{\int y \nu(dy)} \right] = 0, \quad \nu(h) > 0, \quad \text{a.s.} \tag{43}$$

Using Corollary 1 again, in the pointwise sense we have

$$\mathcal{I}_{0,h} \leq_{h''} \mathcal{I}_{0,h''}, \quad \mathcal{I}_{0,h''}(h'') \leq \mathcal{I}_{0,h}(h),$$

implying that in the pointwise sense (since  $\nu$  is a pointwise weak limit of  $\mathcal{I}_{0,h''}$ )

$$\mathcal{I}_{0,h} \leq_{h-} \nu, \quad \nu(h) \leq \mathcal{I}_{0,h}. \tag{44}$$

On the other hand, by assumption  $\mathcal{I}_{0,h} > 0$  a.s., so using Corollary 4-3, we have

$$\mathbb{E} \left[ \ln \frac{h(1-\beta)}{\int y \mathcal{I}_{0,h}(dy)} \right] = 0.$$

The above display together with (43) and (44) implies that

$$\nu = \mathcal{I}_{0,h} \text{ pointwise.} \tag{45}$$

Therefore we have proved that  $\mathcal{I}_{0,h''}$  converges pointwise to the weak limit  $\mathcal{I}_{0,h}$  as  $h'' \rightarrow h$ .

Now, taking into account (40), for any continuous function  $f$  we have

$$\begin{aligned} \int f(x) P_n''(dx) &\xrightarrow{d} \int f(x) \mathcal{I}_{0,h''}(dx) \\ &\xrightarrow[h'' \rightarrow h]{\text{pointwise}} \int f(x) \mathcal{I}_{0,h}(dx) \xleftarrow{d} \int f(x) P_n'(dx). \end{aligned} \tag{46}$$

Note that using (41), for any bounded continuous increasing function  $g$  we have

$$\int g(x) P_n''(dx) \leq \int g(x) P_n(dx) \leq \int g(x) P_n'(dx).$$

Together with (46), this yields

$$\int g(x) P_n(dx) \xrightarrow{d} \int g(x) \mathcal{I}_{0,h}(dx).$$

Since, by (41),  $P_n' \leq_{h-} P_n$  pointwise for any  $n$ , the above display implies that  $P_n$  converges weakly to  $\mathcal{I}_{0,h}$ , which is the same as the weak limit of  $(P_n')$ .

If  $h = S_Q$ , we follow the same procedure, except that to prove (45) we require Corollary 5. □

#### 4.8. Proof of Theorem 4

First we prove two lemmas. Recall the definition of  $S_u$  for  $u \in M_1$ .

**Lemma 5.**  $S_{(\cdot)}$  is a continuous (hence measurable) function on  $M_1$  with the topology of the weak convergence.

*Proof.* Assume that a sequence  $(u_n)$  converges weakly to  $u$ . If  $S_{u_n}$  does not converge to  $S_u$ , then there exists a subsequence  $(u_{n_k})$  such that  $S_{u_{n_k}}$  converges to a limit  $a$  with  $a < S_u$  or  $a > S_u$ . Without loss of generality, assume  $a < S_u$ . We take a positive and continuous function  $f$  supported on

$$\left( \frac{a + S_u}{2}, S_u \right],$$

and then  $\int f(x)u(dx) > 0$ . But  $\int f(x)u_{n_k}(dx)$  converges to 0. This contradicts the weak convergence, which completes the proof.  $\square$

The next lemma generalises Corollary 5.

**Lemma 6.** *For any invariant measure  $\nu$  with  $S_\nu = h$  a.s., we have  $\nu \stackrel{d}{=} \mathcal{I}$ .*

*Proof.* Let  $(P_n)$  be the forward sequence in the random model with  $P_0 \stackrel{d}{=} \nu$  and  $P_0$  independent of  $(\beta_n)$ . By Theorem 2, conditionally on  $P_0$ ,  $P_n$  converges in distribution to the same random measure  $\mathcal{I}$ . Then, unconditionally,  $P_n \stackrel{d}{=} \nu$  converges in distribution to  $\mathcal{I}$ , implying  $\nu \stackrel{d}{=} \mathcal{I}$ .  $\square$

*Proof of Theorem 4.* Let  $\nu$  be any invariant measure. By (10),  $S_\nu \in [S_Q, 1]$ , a.s. By Lemma 5,  $S_\nu$  is a random variable, and then by Theorem 5.3 in [15], there exists a regular conditional distribution of  $\nu$  on  $S_\nu$ .

Conditioning on  $S_\nu$  for both sides of (10), we see that  $(\nu|_{S_\nu})$  must be an invariant measure a.s. By Lemma 6, conditionally on  $S_\nu$ , we have  $(\nu|_{S_\nu}) \stackrel{d}{=} \mathcal{I}$  a.s., where  $\mathcal{I}$  is the random probability measure with parameters  $\mathcal{L}, Q, h = S_\nu$  and satisfies  $\mathbb{P}(S_{\mathcal{I}} = S_\nu | S_\nu) = 1$  a.s. This finishes the proof.  $\square$

### Appendix A. Analysis of $\ln \frac{h(1-b)}{\int x\mathcal{K}_Q(dx)}$ in Kingman’s model

We discuss respectively Theorem 1-1, i.e.

$$\int \frac{Q(dx)}{1-x/h} \geq b^{-1},$$

and Theorem 1-2, i.e.

$$\int \frac{Q(dx)}{1-x/h} < b^{-1}.$$

For the former, let us first compute  $\int x\mathcal{K}_Q(dx)$ :

$$\begin{aligned} \int x\mathcal{K}_Q(dx) &= \int \frac{b\theta_b x Q(dx)}{\theta_b - (1-b)x} \\ &= \int \frac{b\theta_b(x - \theta_b/(1-b))Q(dx) + b\theta_b^2/(1-b)Q(dx)}{\theta_b - (1-b)x} \\ &= \frac{b\theta_b}{1-b} + \frac{b\theta_b^2}{1-b} \int \frac{Q(dx)}{\theta_b - (1-b)x} \\ &= \theta_b, \end{aligned} \tag{47}$$

where the last equality is due to the fact that  $\theta_b$  is the solution of the equation (4). The equation (4) also implies

$$\int \frac{\theta_b Q(dx)}{\theta_b - (1-b)x} = \int \frac{Q(dx)}{1 - (1-b)x/\theta_b} = b^{-1}.$$

Recall that  $\int \frac{Q(dx)}{1-x/h} \geq b^{-1}$ . Then the above display implies that

$$\frac{1}{h} \geq \frac{1-b}{\theta_b}.$$

Taking into account (47), we arrive at

$$\frac{h(1-b)}{\int x\mathcal{K}_Q(dx)} \leq 1, \quad \text{or equivalently, } \ln \frac{h(1-b)}{\int x\mathcal{K}_Q(dx)} \leq 0.$$

Equality holds if and only if

$$\int \frac{Q(dx)}{1-x/h} = b^{-1}.$$

For Theorem 1-2, we have

$$\begin{aligned} \int x\mathcal{K}_Q(dx) &= \int \frac{bxQ(dx)}{1-x/S_Q} + \left(1 - \int \frac{bQ(dx)}{1-x/S_Q}\right) S_Q \\ &= S_Q + \int \frac{b(x-S_Q)Q(dx)}{1-x/S_Q} \\ &= (1-b)S_Q. \end{aligned}$$

Thus we obtain

$$\ln \frac{h(1-b)}{\int x\mathcal{K}_Q(dx)} = \ln \frac{h}{S_Q} \geq 0,$$

where equality holds if and only if  $h = S_Q$ .

In conclusion, if  $h > S_Q$ , then

$$\ln \frac{h(1-b)}{\int x\mathcal{K}_Q(dx)} \leq 0$$

is equivalent to

$$\int \frac{Q(dx)}{1-x/h} \geq b^{-1}$$

(non-condensation case), and

$$\ln \frac{(1-b)S_Q}{\int x\mathcal{K}_Q(dx)} > 0$$

is equivalent to

$$\int \frac{Q(dx)}{1-x/h} < b^{-1}$$

(condensation case).

If  $h = S_Q$ , then

$$\ln \frac{S_Q(1-b)}{\int x\mathcal{K}_Q(dx)} = 0$$

is equivalent to either

$$\int \frac{Q(dx)}{1-x/S_Q} = b^{-1}$$

(non-condensation case) or

$$\int \frac{Q(dx)}{1-x/S_Q} < b^{-1}$$

(condensation case), and

$$\ln \frac{S_Q(1-b)}{\int x\mathcal{K}_Q(dx)} < 0$$

is equivalent to

$$\int \frac{Q(dx)}{1-x/S_Q} > b^{-1}$$

(non-condensation case). The case

$$\ln \frac{S_Q(1-b)}{\int x\mathcal{K}_Q(dx)} > 0$$

does not occur, which is in line with Remark 13.

Therefore, if  $h = S_Q$ , knowing only

$$\ln \frac{S_Q(1-b)}{\int x\mathcal{K}_Q(dx)} = 0$$

does not allow one to determine whether condensation occurs or not.

### Acknowledgements

The author would like to thank Takis Konstantopoulos, Götz Kersting, and Pascal Grange for discussions. The author thanks the anonymous referees for their comments, which greatly improved the presentation of the paper, and for suggesting the intuition for Theorem 3.

### Funding Information

The author acknowledges the support of the National Natural Science Foundation of China (Youth Programme, Grant 11801458).

### Competing Interests

There were no competing interests to declare which arose during the preparation or publication process for this article.

### References

- [1] ALDOUS, D. J. AND BANDYOPADHYAY, A. (2005). A survey of max-type recursive distributional equations. *Ann. Probab.* **15**, 1047–1110.
- [2] BETZ, V., DEREICH, S. AND MÖRTERS, P. (2018). The shape of the emerging condensate in effective models of condensation. *Ann. Inst. H. Poincaré Prob. Statist.* **19**, 1869–1889.



- [3] BÜRGER, R. (1986). On the maintenance of genetic variation: global analysis of Kimurabs continuum-of-alleles model. *J. Math. Biol.* **24**, 341–351.
- [4] BÜRGER, R. (1989). Mutation–selection balance and continuum-of-alleles models. *Math. Biosci.* **12**, 67–83.
- [5] BÜRGER, R. (1998). Mathematical properties of mutation–selection models. *Genetica* **102**, 279–298.
- [6] BÜRGER, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. John Wiley, Chichester, New York.
- [7] CROW, J. F. AND KIMURA, M. (1970). *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- [8] DEREICH, S. AND MÖRTERS, P. (2013). Emergence of condensation in Kingman's model of selection and mutation. *Acta Appl. Math.* **127**, 17–26.
- [9] DIACONIS, P. AND FREEDMAN, D. (1999). Iterated random functions. *SIAM Rev.* **41**, 45–76.
- [10] EVANS, S. N., STEINSALTZ, D. AND WACHTER, K. W. (2013). *A Mutation–Selection Model with Recombination for General Genotypes*. American Mathematical Society, Providence, RI.
- [11] EWENS, W. J. (1979). *Mathematical Population Genetics*. Springer, New York.
- [12] GONZALEZ-CASANOVA, A., KURT, N., WAKOLBINGER, A. AND YUAN, L. (2016). An individual-based model for the Lenski experiment, and the deceleration of the relative fitness. *Stoch. Process. Appl.* **126**, 2211–2252.
- [13] GRANGE, P. (2019). Steady states in a non-conserving zero-range process with extensive rates as a model for the balance of selection and mutation. *J. Phys. A* **52**, 365601.
- [14] HALDANE, J. B. S. (1937). The effect of variation on fitness. *Amer. Naturalist* **71**, 337–349.
- [15] KALLENBERG, O. (1997). *Foundations of Modern Probability*. Springer, Cham.
- [16] KALLENBERG, O. (2017). *Random Measures: Theory and Applications*. Springer, Cham.
- [17] KIMURA, M. (1965). A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proc. Nat. Acad. Sci. USA* **54**, 731–736.
- [18] KINGMAN, J. F. C. (1977). On the properties of bilinear models for the balance between mutation and selection. *Proc. Camb. Phil. Soc.* **80**, 443–453.
- [19] KINGMAN, J. F. C. (1978). A simple model for the balance between selection and mutation. *J. Appl. Prob.* **15**, 1–12.
- [20] KINGMAN, J. F. C. (1980). *Mathematics of Genetic Diversity*. Society for Industrial and Applied Mathematics, Philadelphia.
- [21] STEINSALTZ, D., EVANS, S. N. AND WACHTER, K. W. (2005). A generalized model of mutation–selection balance with applications to aging. *Adv. Appl. Math.* **35**, 16–33.
- [22] YUAN, L. (2017). A generalization of Kingman's model of selection and mutation and the Lenski experiment. *Math. Biosci.* **285**, 61–67.
- [23] YUAN, L. (2020). Kingman's model with random mutation probabilities: convergence and condensation II. *J. Statist. Phys.* **181**, 870–896.