ALEA, Lat. Am. J. Probab. Math. Stat. 20, 1041–1075 (2023)
DOI: 10.30757/ALEA.v20-38

An evolution model with uncountably many alleles

Daniela Bertacchi, Jüri Lember and Fabio Zucca

Dipartimento di Matematica e Applicazioni, Università di Milano–Bicocca, via Cozzi 53, 20125 Milano, Italy
E-mail address: daniela.bertacchi@unimib.it

Institute of Mathematics and Statistics, University of Tartu, J. Liiv 2, 50409 Tartu, Estonia.
E-mail address: juri.lember@ut.ee

Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy.
E-mail address: fabio.zucca@polimi.it

Abstract. We study a class of evolution models, where the breeding process involves an arbitrary exchangeable process, allowing for mutations to appear. The population size n is fixed, hence after breeding, selection is applied. Individuals are characterized by their genome, picked inside a set $\mathcal{X}$ (which may be uncountable), and there is a fitness associated to each genome. Being less fit implies a higher chance of being discarded in the selection process. The stationary distribution of the process can be described and studied. We are interested in the asymptotic behavior of this stationary distribution as n goes to infinity. Choosing a parameter $\lambda > 0$ to tune the scaling of the fitness when n grows, we prove limiting theorems both for the case when the breeding process does not depend on n, and for the case when it is given by a Dirichlet process prior. In both cases, the limit exhibits phase transitions depending on the parameter λ.

1. Introduction

The model and setup. We study the (uncountable) infinite alleles evolution model, where the breeding and mutation is governed by an $\mathcal{X}$-valued infinitely exchangeable process $\xi_1, \xi_2, \ldots$, we shall refer to ξ as the breeding process. Throughout the paper $\mathcal{X}$ stands for the Polish (i.e. complete and separable metric) space of all alleles, the elements of $\mathcal{X}$ are called (geno)types in sequel. By the de Finetti-Hewitt-Savage theorem, the process ξ is in one-to-one correspondence with a probability measure π on the Borel σ-algebra of all probability measures $\mathcal{P}$ equipped with the topology of weak convergence on $\mathcal{X}$, because for every $n \in \mathbb{N}$, and $A_i \in \mathcal{B}(\mathcal{X})$, $i = 1, \ldots, n$,

$$\mathbf{P}(\xi_1 \in A_1, \ldots, \xi_n \in A_n) = \int_{\mathcal{P}} \prod_{i=1}^{n} q(A_i)\pi(\mathrm{d}q), \tag{1.1}$$

Received by the editors September 27th, 2022; accepted July 18th, 2023.

2010 *Mathematics Subject Classification.* 60J05, 60B10, 92D15.

Key words and phrases. Population genetics, Moran model, Dirichlet process, large population limit.

Daniela Bertacchi and Fabio Zucca have been supported by GNAMPA-INdAM. Jüri Lember has been supported by Estonian Research Council grant PRG865.

where $\mathcal{B}(\mathcal{X})$ stands for Borel σ-algebra of $\mathcal{X}$ (see, e.g. Hjort et al., 2010, Ch 3 or Fortini et al., 2000, Theorem 1). Hence we identify the process ξ with π (see Subsection 2 for details). The measure π will be referred to as the prior measure. In models of evolution, a commonly used prior is the law of the Dirichlet process $DP(m, \bar{\alpha})$, where $\bar{\alpha}$ is a (typically non-atomic) probability measure on $\mathcal{X}$ and $m > 0$ is so-called concentration or precision parameter. With this prior the breeding process is the following: for every $n \geq 1$, $A \in \mathcal{B}(\mathcal{X})$, and population $x_1, \dots, x_n$

$$P\big(\xi_{n+1} \in A | \xi_1 = x_1, \dots, \xi_n = x_n\big) = \frac{m}{m+n}\bar{\alpha}(A) + \frac{n}{m+n} \cdot \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}(A) \tag{1.2}$$

where $P(\xi_1 \in A) = \bar{\alpha}(A)$. If $x_1^*, \dots, x_k^*$ are the distinct values of $x_1, \dots, x_n$ with $n_1, \dots, n_k$ being their frequencies, the conditional distribution above can be interpreted as follows:

$$\xi_{n+1} | \xi_1 = x_1, \dots, \xi_n = x_n \sim \begin{cases} \bar{\alpha}, & \text{with probability } \frac{m}{m+n}; \\ \delta_{x_j^*}, & \text{with probability } \frac{n_j}{m+n} \quad j = 1, \dots, k. \end{cases}$$

(see, e.g. Ghosal and van der Vaart, 2017; Hjort et al., 2010). This interpretation allows to obtain the sequence $\xi_1, \xi_2, \dots$ by a very simple procedure, known as the generalized Polya urn scheme. The scheme is very easy to implement making Dirichlet process priors popular in applications (see for instance samplers in Ghosal and van der Vaart, 2017, Ch. 4.5). When $\bar{\alpha}$ is non-atomic, then with probability $m/(m+n)$ the random variable ξ_{n+1} takes a new value that is not previously seen in $x_1, \dots, x_n$ – a mutation. Hence the ratio $m/(m+n)$ can be interpreted as the mutation probability. In the literature of evolution models, the Polya urn scheme with non-atomic $\bar{\alpha}$ is often referred to as Hoppe urn (with m typically denoted by θ), the only difference between the two urns is that in the Polya urn the mutations are labeled and $\bar{\alpha}$ specifies their origin. In particular, the celebrated Ewens sampling formula, along with its consequences, holds under Polya urn scheme, and therefore the Dirichlet process prior is central in evolution theory.

We start with a fixed population size n. The process ξ models the breeding: given the population $x_1, \dots, x_n$, the new genotype x_{n+1} is bred from the conditional distribution of ξ_{n+1} given $\xi_i = x_i$, $i = 1, \dots, n$, denoted as $P_\xi(\cdot | x_1, \dots, x_n)$. Observe that since the order in the population is arbitrary, exchangeability is a natural assumption about the breeding process ξ. After a new individual with genotype x_{n+1} is born, it either replaces an already existing member of the population, or it is discarded, and the population remains unchanged. The probability that x_{n+1} is kept in the population depends on the fitnesses of all population members. So, in what follows, let $w : \mathcal{X} \to \mathbb{R}^+$ be a bounded continuous strictly positive function, assigning a fitness to every type. The bigger $w(x_{n+1})$, the more likely that a newborn member replaces an already existing one in the population. There are several selection schemes possible. In Lember and Watkins (2022), the following schemes were introduced.

- **Single tournament selection:**
  1. Sample $x_{n+1} \sim P_\xi(\cdot \mid x_1, \dots, x_n)$
  2. Sample i randomly from $\{1, \dots, n\}$
  3. With probability $\frac{w(x_{n+1})}{w(x_i) + w(x_{n+1})}$ replace x_i with x_{n+1} and discard x_i, otherwise discard x_{n+1}.
- **Inverse fitness selection:**
  1. Sample $x_{n+1} \sim P_\xi(\cdot \mid x_1, \dots, x_n)$
  2. Sample i from $\{1, \dots, n+1\}$ with probabilities proportional to $\{\frac{1}{w(x_1)}, \dots, \frac{1}{w(x_{n+1})}\}$
  3. If $i < n + 1$, then replace x_i by x_{n+1}.

Both selection schemes define a Markov kernel on $\mathcal{X}^n$. Lemma 2.1 below proves that both kernels satisfy the detailed balance equation with stationary measure

$$P_n(A) := \frac{1}{Z_n} \int_A \prod_{j=1}^n w(x_j) P_\xi^n(\mathrm{d}\mathbf{x}), \quad A \in \mathcal{B}(\mathcal{X}^n), \tag{1.3}$$

where P_ξ^n is the law of $(\xi_1, \dots, \xi_n)$, $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ and Z_n is the normalizing constant. Hence the stationary (or limit) distribution of the genotypes in n-elemental population has clear and explicit closed form, depending solely on w and π.

The measure P_n in (1.3) is the main object of interest. The article focuses on the limit of P_n when the population size n grows and the fitness function w_n and prior π_n both might depend on n. Since P_n is defined on different domains $\mathcal{X}^n$ the definition of a limit is not straightforward. To overcome that problem, we consider two parallel approaches:

- The first approach is to consider the triangular array of random variables
$$(X_{1,n}, \dots, X_{n,n}) \sim P_n, \quad n = 1, 2, \dots$$
and ask: is there a limit stochastic process $X_1, X_2, \dots$ such that for every m and for every m-tuple of integers $t_1, t_2, \dots, t_m$, it holds (as $n \to \infty$) that
$$(X_{t_1,n}, X_{t_2,n}, \dots, X_{t_m,n}) \Rightarrow (X_{t_1}, X_{t_2}, \dots, X_{t_m}). \tag{1.4}$$
If such a limit process exists, it can be considered as an approximation of the population $(X_{1,n}, \dots, X_{n,n})$ for big n. Theorem 3.1 provides the main technical tool for proving the convergence (1.4) and, hence, the existence of the limit process.
- The second approach is to transfer the measures P_n into the measures Q_n on $\mathcal{P}$. For that we define the mapping g that maps a vector $\mathbf{x}$ to its empirical measure:
$$g : \mathcal{X}^n \to \mathcal{P}, \quad g(\mathbf{x}) = \frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n}) \tag{1.5}$$
and we define Q_n as $P_n g^{-1}$, i.e. the pushforward measure
$$Q_n(E) := P_n\big(g^{-1}(E)\big), \quad E \in \mathcal{B}(\mathcal{P}). \tag{1.6}$$
In other words, Q_n is the distribution of $g(X_1, \dots, X_n)$, where $(X_1, \dots, X_n) \sim P_n$. Since P_n is invariant with respect to permutations, i.e. the n-dimensional random vector having distribution as P_n is exchangeable, we see that P_n can be uniquely restored from Q_n, so in a sense they are the same. In Subsection 2.2, we shall argue that g is measurable, i.e. Q_n is well defined. Since the measures Q_n are defined on the same domain $\mathcal{P}$, the question now is the existence of a limit measure Q^* such that $Q_n \Rightarrow Q^*$ (the weak convergence). The Q_n-counterpart of Theorem 3.1 is Theorem 3.4 that provides necessary conditions in terms of w_n and π_n for existence of Q^*. Since the proof of Theorem 3.4 is based on large deviation inequality, an additional assumption that $\mathcal{X}$ is compact is imposed.

The phase transition results. For the first convergence results in Section 4 we consider the case where the prior is arbitrary and independent of n, $\pi_n =: \pi$. As weight functions, we take

$$w_n(x) = \exp\left[-\frac{\phi(x)}{n^\lambda}\right], \tag{1.7}$$

where $\lambda \geq 0$ and the function $\phi(x)$ is nonnegative, continuous and bounded. The parameter λ controls how fast the differences between fitness functions w_n vary when n increases. The case $\lambda = 0$ corresponds to the case $w_n = w$ for every n. Observe that P_n in (1.3) is invariant with respect to multiplying w by a positive constant, hence when w is bounded from above, there is no loss of generality in taking it to be bounded by 1 as (1.7) implies. Let us note that often, in the two-allele model ($\mathcal{X} = \{a, A\}$), the fitness of the two alleles a and A is taken as 1 and $(1 + \gamma/n)$

respectively. Since $(1+\gamma/n) \approx \exp[\gamma/n]$, we retrieve (approximately) the fitness in (1.7) by choosing $\phi(a) = 0$, $\phi(A) = \gamma$ and $\lambda = 1$. We shall see, from the limiting results as n goes to infinity, that $\lambda = 1$ is indeed, in a sense, the right scaling.

Theorems 4.2 and 4.7 are the main phase transition theorems for arbitrary π. The results of both theorems can summarized as follows:

- **The case $\lambda > 1$.** Then (1.4) holds with the limit process being equal to the breeding process ξ and $Q_n \Rightarrow \pi$. This means that when $\lambda > 1$, then the differences between fitnesses vanish so quickly that the selection has no influence in the limit.
- **The case $\lambda = 1$.** Then (1.4) holds with the limit process being an infinitely exchangeable process with prior measure $\bar{\pi}$ (specified in Theorem 4.2) that depends on ϕ and π and is different from π. Then also $Q_n \Rightarrow \bar{\pi}$.
- **The case $\lambda \in [0, 1)$.** In this case we impose an additional mild assumption on ϕ (that in particular guarantees the uniquess of the minimum x_o), and we also assume that the support of π contains δ_{x_o}. Then (1.4) holds with the limit process being degenerate with one possible path $x_o, x_o, \cdots$ and $Q_n \Rightarrow \delta_{q^*}$, where $q^* = \delta_{x_o}$. This means that when $\lambda < 1$ then the selection is so strong that breeding has no influence in the limit and only the fittest type x_o (that maximizes w_n for every n) survives.

In Section 5 we consider the case when the prior measure is the law of $\mathrm{DP}(m_n, \bar{\alpha})$, but the fitness function is still (1.7). We let the concentration parameter m_n scale with n, depending on the same parameter λ as follows: $m_n = cn^{1-\lambda}$, where $c > 0$. This choice of m_n, as we see in the following, leads to interesting limiting behaviours, while keeping the number of free parameters at a minimum. When $\lambda = 1$, then $m_n = c$, and therefore π is independent of n, hence this case is the case considered above. However, the case $\lambda \in [0, 1)$ needs special treatment. Observe that when $\lambda = 0$, then $w_n = w$ and the mutation probability $m_n/(n + m_n) = c/(1 + c)$ is independent of the population size n and this makes the case $\lambda = 0$ appealing.

The results with Dirichlet process prior, Theorems 5.3, 5.7 and 5.9 can be summarized as follows, the additional assumptions are that $\mathcal{X}$ is compact, $\bar{\alpha}$ has full support and x_o is the unique minimizer of ϕ. We remark that in the case $\lambda \in (0, 1)$, the limit measure is independent of λ.

- **The case $\lambda = 0$.** Then (1.4) holds with the limit process being an iid process with $X_i \sim r^*$ and $Q_n \Rightarrow \delta_{r^*}$. The measure r^* depends on the inequality

$$\int_{\mathcal{X}} \frac{w(x_o)}{w(x_o) - w(x)} \bar{\alpha}(\mathrm{d}x) \geq \frac{1 + c}{c}. \tag{1.8}$$

When (1.8) holds, then r^* has density $r^*(x)$ with respect to $\bar{\alpha}$:

$$r^*(x) = \frac{cw(x)}{\theta(1 + c) - w(x)},$$

where $\theta > 0$ depends on w, $\bar{\alpha}$ and c (see Lemma 5.1). Observe that the density is with respect to $\bar{\alpha}$-measure, so when $\bar{\alpha}$ has an atom, then r^* (i.e. the limit population) has the same atom, but its mass is re-weighted. But when (1.8) fails, then r^* has an absolutely continuous part (with respect to $\bar{\alpha}$) with density

$$r^*_a(x) = \frac{cw(x)}{w(x_o) - w(x)}$$

but also an atom at x_o. When (1.8) fails, then $\bar{\alpha}$ cannot have an atom at x_o, but w is peaked so heavily in neighborhood of x_o that in the limit measure an atom appears. Thus, when (1.8) fails, then in the limit population there is a fixed proportion of individuals with the fittest type x_o.

- **The case** $\lambda \in (0,1)$**.** Then (1.4) holds with the limit process being an iid process with $X_i \sim q^*$ and $Q_n \Rightarrow q^*$. The measure q^* depends on the inequality

$$\int_{\mathcal{X}} \frac{1}{\phi(x) - \phi(x_o)} \bar{\alpha}(\mathrm{d}x) \geq \frac{1}{c}. \tag{1.9}$$

When (1.9) holds, then q^* has density $q^*(x)$ with respect to $\bar{\alpha}$:

$$q^*(x) = \frac{c}{\phi(x) - c - \theta},$$

where $\theta > 0$ is a parameter. When (1.9) fails, then q^* has an absolutely continuous part (with respect to $\bar{\alpha}$) with density

$$q^*_a(x) = \frac{c}{\phi(x) - \phi(x_o)}$$

but also an atom at x_o. Thus, when (1.9) fails, hence $\bar{\alpha}(x_o) = 0$, in the limit an atom at x_o is created.
- **The case** $\lambda = 1$**.** This is a special case of the constant prior with $\pi = Dir(c, \bar{\alpha})$. Thus (1.4) holds with the limit process being an infinitely exchangeable process with prior measure $\bar{\pi}$ (specified in Theorem 4.2) and $Q_n \Rightarrow \bar{\pi}$.
- **The case** $\lambda > 1$**.** Here the influence of fitness vanishes and (1.4) holds with the limit process being degenerate with one possible path $X, X, \cdots$ with $X \sim \bar{\alpha}$ and $Q_n \Rightarrow \delta_X$, where $X \sim \bar{\alpha}$.

The case of finite $\mathcal{X}$ *and the relation with previous work.* The literature on mathematical population genetics is vast and focuses on various aspects of the evolution of traits within a population (see Ewens, 2004 and references therein). A common feature of these models is the fact that individuals, characterized by their genome (or "type", "trait"), breed and die. Mutations may occur and a fitness be associated to each type. The population can be modelled as having a varying or a fixed size. In the first case the process is usually a birth-death process and one can focus either on the equilibrium when time grows, or on the trajectories. For instance the asymptotic distribution of fitnesses is studied by Guiol et al. (2011); Ben-Ari et al. (2011); Bertacchi et al. (2018) in an evolution scheme where a random number of least fit individuals die at each generation, while Bertacchi et al. (2016); Iwasa and Levin (1995) study the effect of random/deterministic events on this distribution; Bansaye et al. (2019) and Berzunza et al. (2021) study the convergence of the evolutionary process as the population size goes to infinity. An example of a model where the fitness function has more than one local maximum is the NK-model, which has been rigorously studied in Durrett and Limic (2003) and Evans and Steinsaltz (2002). The population is assumed to have a constant size in classical models such as the Wright-Fisher and the Moran models, but even with this assumptions there are still many theoretical challenges and applications (see for instance Durrett and Mayberry, 2011; Schweinsberg, 2017a,b).

In our model, the population has constant size n, breeding is conditional sampling from a very general exchangeable process and selection takes place at death (where less fit individuals are more likely to be removed). We are interested in the asymptotic behaviour, as the size of the population increases, of the stationary distribution of types.

The case of finitely many types $|\mathcal{X}| = K < \infty$ was treated in Lember and Watkins (2022). Then $\mathcal{P}$ is just a simplex and when π is the Dirichlet distribution $\mathrm{Dir}(\alpha_1, \ldots, \alpha_K)$, then the breeding process can be considered as a version of the well known Moran model without selection (see Lember and Watkins, 2022, Section 2.1). When the selection (either single tournament or inverse fitness) is added, then we end up with a version of the Moran model with breeding and selection. The stationary measure P_n in this case (π equals to Dirichlet distribution) in terms of allele counts is as follows

$$P_n(n_1,\ldots,n_K)=\frac{1}{Z_n}\frac{n!}{n_1!\cdots n_K!}\frac{(\alpha_1)_{n_1}\cdots(\alpha_K)_{n_K}}{(|\alpha|)_n}w^{n_1}(1)\cdots w^{n_K}(K), \tag{1.10}$$

where $(\alpha)_n=\alpha(\alpha+1)\cdots(a+n-1)$, $|\alpha|=\alpha_1+\cdots+\alpha_K$ and $n_k\geq 0$ stands for the number of type k in $(x_1,\ldots,x_n)$, thus $n_1+\cdots+n_K=n$. Since P_n is exchangeable, it can be equally presented in terms of counts and in the literature it is typically done so. As pointed out in Lember and Watkins (2022) there are many other versions of Moran models leading the same stationary distribution (the parameters α are obtained then from mutation probabilities). Hence the case $\pi=\mathrm{Dir}(\alpha_1,\ldots,\alpha_K)$ is an important special case also in the finite allele model.

Although one may argue that in reality the set of types $\mathcal{X}$ is always finite, we emphasize that its cardinality is much larger than the population size itself. Thus, since we are considering limits when the population size grows to infinity, it is reasonable to assume that $\mathcal{X}$ is infinite, which poses some difficulties in studying this model. For finite $\mathcal{X}$, the limit theorems of this paper hold true as well (they are just a special case), but the proofs are much simpler. Hence in a sense the current article can be considered as a generalization of Lember and Watkins (2022), but the generalization is far from being trivial. For general $\mathcal{X}$, a new machinery needs to be built, and it is the purpose of the current paper. The difference between general and finite $\mathcal{X}$ is well illustrated by the Dirichlet process results (Theorems 5.3 and 5.7). In the finite case, since $\alpha_k>0$ for every $k=1,\ldots,K$, the inequalities (1.8) and (1.9) both hold. In the finite case the limit probability measures r^* and q^* are elements of a simplex (thus K-dimensional vectors) as follows:

$$r^*(k)=\frac{w(k)\alpha_k}{\theta(1+|\alpha|)-w(k)},\quad q^*(k)=\frac{\alpha_k}{\phi(k)+|\alpha|-\theta},\quad k=1,\ldots,K$$

where in both formulas θ is a parameter. Since in the finite case $c=|\alpha|$ and $\bar{\alpha}_k=\frac{\alpha_k}{|\alpha|}$, we see that these measures are indeed the same as given by Theorems 5.3 and 5.7. However, quite surprisingly for general $\mathcal{X}$ the additional atom appears. This is something one cannot predict based solely on the results of Lember and Watkins (2022). Also the proofs of Theorems 4.2 and 4.7 for general $\mathcal{X}$ are essentially different from the ones in the case of finite $\mathcal{X}$, they are based on large deviation results and therefore the additional assumption of compactness is needed. We also would like to stress that all limits in Lember and Watkins (2022) as well as all limits in the in the current paper are obtained without diffusion approximation. For example, when $\mathcal{X}$ is finite, π is Dirichlet distribution and $\lambda=1$, the the limit density – sometimes called as Wright's formula – can be found in the literature (for references, see Lember and Watkins, 2022, Section 3.4), typically connected to the diffusion approximation. However, while the proof in Lember and Watkins (2022) uses fairly simple mathematics and no diffusion approximation, the generalization to general $\mathcal{X}$ (Theorem 4.2 in the current article) uses more involved mathematics, but again no diffusion approximation.

Outline of the paper. In Section 2, the model and the main objects of the article, are formally defined. In Subsection 2.1, the detailed balance equation is proven showing that P_n is indeed the stationary measure of the model (with population size n). In Subsection 2.2 we give an alternative representation of P_n and define the measure Q_n. In Section 3, the limit process (infinite population) and the sense of convergence are defined. The main results of Section 3 are Theorems 3.1 and 3.4. The first of them proves the existence of the limit process under rather general assumptions and the second theorem shows there also exists a limit measure Q such that $Q_n\Rightarrow Q$ (under the additional assumption that $\mathcal{X}$ is compact). These theorems are the basis of the paper. Section 4 is devoted to the case when the prior measure π is arbitrary but independent of n. The main results of that section are Theorems 4.2 and 4.7; these two theorems together give the first phase transition result as described in Introduction. In Section 5, the Dirichlet process prior is considered. The main results are Theorem 5.3, 5.7 and 5.9 which provide phase transition results for that case. The proofs of these theorems are rather technical and therefore they are presented in Section 6.

2. Preliminaries

Recall that $\mathcal{P}$ stands for the set of all probability measures on Borel σ-algebra $\mathcal{B}(\mathcal{X})$. In what follows, we shall denote the elements of $\mathcal{P}$ by q. For any integrable function f on $\mathcal{X}$, we shall denote by

$$\langle f, q\rangle := \int_{\mathcal{X}} f(x)q(\mathrm{d}x).$$

The set $\mathcal{P}$ is equipped with Prokhorov metric and so this is a complete separable metric space as well, see (Billingsley, 1999, p. 72). Prokhorov metric metrizes the weak convergence of probability measures, denoted by $q_n \Rightarrow q$ in the sequel, and the Borel σ-algebra $\mathcal{B}(\mathcal{P})$ is such that for any $A \subset \mathcal{B}(\mathcal{X})$, the mapping $q \mapsto q(A)$ is $\mathcal{B}(\mathcal{P})$-measurable (see e.g. Ghosal and van der Vaart (2017, Proposition A.5)). Also for any continuous bounded function f on $\mathcal{X}$, the function $q \mapsto \langle f, q\rangle$ is continuous and hence $\mathcal{B}(\mathcal{P})$-measurable as well. In what follows, we shall also see n-fold product measures q^n on $\mathcal{B}(\mathcal{X}^n)$. Since $q \mapsto q(A)$ is measurable for every $A \in \mathcal{B}(\mathcal{X})$, then also $q \mapsto q^n(A)$ is measurable for every $A \in \mathcal{B}(\mathcal{X}^n)$. This follows from Dynkin's $\pi - \lambda$ theorem: clearly for any cylinder $A = A_1 \times \cdots \times A_n$ the mapping $q \mapsto q^n(A) = q(A_1)\cdots q(A_n)$ is measurable (as a product of measurable functions). The set $\Lambda = \{A \in \mathcal{B}(\mathcal{X}^n) : q \mapsto q^n(A) \text{ is measurable }\}$ is a λ-system (i.e. contains $\mathcal{X}^n$ and closed with respect to complements and disjoint unions) containing all cylinders. Since $\mathcal{X}$ is Polish, the cylinders generate $\mathcal{B}(\mathcal{X}^n)$, and so, by Dynkin's $\pi - \lambda$ theorem, $\mathcal{B}(\mathcal{X}^n) \subset \Lambda$.

To see the one-to-one correspondence between breeding processes ξ and measures π, observe that for every measure π, there is a process ξ satisfying equation (1.1). Indeed, for any $A \in \mathcal{B}(\mathcal{X}^n)$, the map $q \mapsto q^n(A)$ is integrable then $P^n_\xi(A) = \int_{\mathcal{X}} q^n(A)\pi(\mathrm{d}q)$ exists and the family $\{P^n_\xi\}$ satisfies Kolmogorov's consistency conditions. The claim follows from Aliprantis and Border (2006, Theorems 12.7 and 15.26). The other direction – to every exchangeable ξ there corresponds a measure π – follows from de Finetti-Hewitt-Savage representation. Recall that $P_\xi(\cdot|\mathbf{x})$ stands for the conditional distribution of ξ_{n+1}. The existence of a regular version for the conditional probability in Polish spaces is a consequence of Pfanzagl and Pierlo (1966, Theorem 7.8). Indeed it is enough that the σ-algebra contains a sub-σ-algebra which is separable (generated by a countable collection of sets) and the probability measure is compact approximable. Both conditions hold for a probability measure on the Borel σ-algebra of a Polish metric space: the first one is trivial and the second one follows from Aliprantis and Border (2006, Theorem 11.20).

2.1. *Detailed balance equation.*

Kernels. Recall the two selection schemes: single tournament and inverse fitness. Both define a Markov chain with uncountable state space $\mathcal{X}^n$. When $\mathcal{X}$ is finite, as in Lember and Watkins (2022), then the corresponding transition matrix is easy to define. We now define the corresponding transition kernels for both schemes. Recall that w is a strictly positive, bounded and continuous fitness function on $\mathcal{X}$.

Let, for every $A \in \mathcal{B}(\mathcal{X}^n)$, $\mathbf{x} \in \mathcal{X}^n$ and $k = 1, \ldots, n$

$$A_k(\mathbf{x}) := \{x \in \mathcal{X} : (\mathbf{x}_1, \ldots, \mathbf{x}_{k-1}, x, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_n) \in A\}.$$

Observe that when $A = A_1 \times \cdots \times A_n$ is a cylindrical set, then

$$A_k(\mathbf{x}) = \begin{cases} A_k, & \text{when } x_j \in A_j, \text{ for every } j \in \{1, \ldots, k-1, k+1, \ldots, n\}; \\ \emptyset, & \text{else.} \end{cases}$$

The transition kernel corresponding to the single tournament selection is

$$P(\mathbf{x},A) =\frac{1}{n}\sum_{k=1}^{n} P_k(\mathbf{x},A), \quad \text{where}$$
$$P_k(\mathbf{x},A) := \int_{A_k(\mathbf{x})} \frac{w(x)}{w(x)+w(x_k)} P_\xi(\mathrm{d}x|\mathbf{x}) + b_k(\mathbf{x})\delta_{x_k}(A_k(\mathbf{x})),$$
$$b_k(\mathbf{x}) := \int_{\mathcal{X}} \frac{w(x_k)}{w(x)+w(x_k)} P_\xi(\mathrm{d}x|\mathbf{x}).$$

Here $b_k(\mathbf{x})$ is the probability that a newborn individual x_{n+1} looses the tournament to x_k. Hence, the first term of $P_k(\mathbf{x},A)$ is the probability that x_{n+1} wins over x_k and is born in $A_k(\mathbf{x})$; the second term is the probability that x_{n+1} looses the tournament to x_k. Clearly $P_k(\mathbf{x},\cdot)$ a probability measure on $\mathcal{B}(\mathcal{X}^n)$. The weight n^{-1} represents the fact that all individuals in population have equal probability to be picked for the tournament. We observe that, applying Dynkin Theorem, one can prove that $x \mapsto P_k(\mathbf{x},A)$ is measurable for every $A \in \mathcal{B}(\mathcal{X}^n)$.

The transition kernel corresponding to the inverse fitness selection is

$$\widetilde{P}(\mathbf{x},A) = \sum_{k=1}^{n} \widetilde{P}_k(\mathbf{x},A) + c(\mathbf{x})\delta_{\mathbf{x}}(A), \quad \text{where}$$
$$\widetilde{P}_k(\mathbf{x},A) := \int_{A_k(\mathbf{x})} \frac{1}{\sum_{j=1}^{n+1} w(x_k)/w(x_j)} P_\xi(\mathrm{d}x_{n+1}|\mathbf{x})$$
$$c(\mathbf{x}) := \int_{\mathcal{X}} \frac{1}{\sum_{j=1}^{n+1} w(x_{n+1})/w(x_j)} P_\xi(\mathrm{d}x_{n+1}|\mathbf{x}).$$

Here $c(\mathbf{x})$ is the probability that x_{n+1} is chosen and so nothing is changed, the first term in $\widetilde{P}_k(\mathbf{x},A)$ is the probability that x_k is chosen and newborn x_{n+1} is in $A_k(\mathbf{x})$.

Reversibility. The following lemma shows the P_n, defined in (1.3), is the stationary measure for both single tournament and inverse fitness kernel, and the stationary process is reversible.

Lemma 2.1. *Let $P(\mathbf{x},A)$ be the transition kernel corresponding to the single tournament selection (resp. to the inverse fitness selection). Then, for every $B, A \in \mathcal{B}(\mathcal{X}^n)$, it holds*

$$\int_B P(\mathbf{x},A)P_n(\mathrm{d}\mathbf{x}) = \int_A P(\mathbf{x},B)P_n(\mathrm{d}\mathbf{x}). \tag{2.1}$$

Proof: It suffices to prove (2.1) if A and B are both cylinders: $A = A_1 \times \cdots \times A_n$, $B = B_1 \times \cdots \times B_n$. Let us consider the single tournament selection kernel. For any fixed k

$$\int_B P_k(\mathbf{x},A)P_n(\mathrm{d}\mathbf{x}) =$$
$$\frac{1}{Z_n}\int_B\int_{A_k(\mathbf{x})} \frac{w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi(\mathrm{d}x_{n+1}|\mathbf{x}) w(x_1)\cdots w(x_n) P_\xi^n(\mathrm{d}\mathbf{x}) + \int_{A\cap B} b_k(\mathbf{x})P_n(\mathrm{d}\mathbf{x}) =$$
$$\frac{1}{Z_n}\int_B\int_{A_k(\mathbf{x})} \frac{w(x_1)\cdots w(x_n)w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}\mathbf{x},\mathrm{d}x_{n+1}) + \int_{A\cap B} b_k(\mathbf{x})P_n(\mathrm{d}\mathbf{x}) = (\$).$$

Now, if we define

$$BA_k := \{\mathbf{x}\in\mathcal{X}^{n+1} \colon (x_1,\ldots,x_n)\in B, (x_1,\ldots,x_{k-1},x_{n+1},x_{k+1},\ldots,x_n)\in A\}$$
$$AB_k := \{\mathbf{x}\in\mathcal{X}^{n+1} \colon (x_1,\ldots,x_n)\in A, (x_1,\ldots,x_{k-1},x_{n+1},x_{k+1},\ldots,x_n)\in B\},$$

we have that one set can be obtained from the other by swapping x_k and x_{n+1}. Whence

$$\int_B \int_{A_k(\mathbf{x})} \frac{w(x_1)\cdots w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}\mathbf{x}, \mathrm{d}x_{n+1}) =$$
$$\int_{BA_k} \frac{w(x_1)\cdots w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}x_1,\ldots,\mathrm{d}x_k,\ldots,\mathrm{d}x_{n+1}) =$$
$$\int_{BA_k} \frac{w(x_1)\cdots w(x_k)\cdots w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}x_1,\ldots,\mathrm{d}x_{n+1},\ldots \mathrm{d}x_k) =$$
$$\int_{AB_k} \frac{w(x_1)\cdots w(x_{n+1})\cdots w(x_k)}{w(x_k)+w(x_{n+1})} P_\xi^{n+1}(\mathrm{d}x_1,\ldots,\mathrm{d}x_k,\ldots \mathrm{d}x_{n+1}) =$$
$$\int_A \int_{B_k(\mathbf{x})} \frac{w(x_1)\cdots w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}\mathbf{x}, \mathrm{d}x_{n+1}).$$

Although the sets BA_k and BA_k are, in general, different, the last equality holds because the function as well as the measure is invariant with respect to change x_{n+1} and x_k. Thus,

$$(\$) = \frac{1}{Z_n}\int_A \int_{B_k(\mathbf{x})} \frac{w(x_1)\cdots w(x_{n+1})}{w(x_{n+1})+w(x_k)} P_\xi^{n+1}(\mathrm{d}\mathbf{x}, \mathrm{d}x_{n+1}) + \int_{A\cap B} b_k(\mathbf{x})P_n(\mathrm{d}\mathbf{x}) = \int_A P_k(\mathbf{x},B)P_n(\mathrm{d}\mathbf{x})$$

and this concludes the first part of the proof.

For the inverse fitness kernel, we proceed similarly. Clearly,

$$\int_B \widetilde{P}(\mathbf{x},A)P_n(\mathrm{d}\mathbf{x}) = \sum_{k=1}^n \int_B \widetilde{P}_k(\mathbf{x},A)P_n(\mathrm{d}\mathbf{x}) + \int_{B\cap A} c(\mathbf{x})P_n(\mathrm{d}\mathbf{x}).$$

Let now $k = 1,\ldots,n$ be fixed and, as previously, we obtain

$$\int_B \widetilde{P}_k(\mathbf{x},A)P_n(\mathrm{d}\mathbf{x}) =$$
$$\frac{1}{Z_n}\int_B \int_{A_k(\mathbf{x})} \frac{w(x_1)\cdots w(x_n)}{\sum_{j=1}^{n+1} w(x_k)/w(x_j)} P_\xi(\mathrm{d}x_{n+1}|\mathbf{x})P_\xi^n(\mathrm{d}\mathbf{x}) =$$
$$\frac{1}{Z_n}\int_{BA_k} \frac{w(x_k)w(x_{n+1})}{\prod_{j=1}^{n+1} w(x_j)} \Big(\sum_{j=1}^{n+1} \frac{1}{w(x_j)}\Big)^{-1} P_\xi^{n+1}(\mathrm{d}x_1,\ldots,\mathrm{d}x_{n+1}) =$$
$$\frac{1}{Z_n}\int_{AB_k} \frac{w(x_k)w(x_{n+1})}{\prod_{j=1}^{n+1} w(x_j)} \Big(\sum_{j=1}^{n+1} \frac{1}{w(x_j)}\Big)^{-1} P_\xi^{n+1}(\mathrm{d}x_1,\ldots,\mathrm{d}x_{n+1}) =$$
$$\int_A \widetilde{P}_k(\mathbf{x},B)P_n(\mathrm{d}\mathbf{x}).$$

□

2.2. *The measures P_n and Q_n.* In the previous section we saw that the measure P_n defined as in (1.3) is a stationary measure for different selection schemes. The main objective of the current article is to study the asymptotic behavior of P_n as the population size n increases. To be more general, we shall assume that the fitness functions w_n and the prior measures π_n depend on n hence, for every n, the measure P_n on $\mathcal{B}(\mathcal{X}^n)$ is the following

$$P_n(A) = \frac{1}{Z_n}\int_A \prod_{j=1}^n w_n(x_j)P_\xi^n(\mathrm{d}\mathbf{x}) = \frac{1}{Z_n}\int_{\mathcal{P}}\int_A \prod_{j=1}^n w_n(x_j)q(\mathrm{d}x_j)\pi_n(\mathrm{d}q). \tag{2.2}$$

The second equality in (2.2) follows from the fact that, for every nonnegative measurable $f : \mathcal{X}^n \to \mathbb{R}^+$, it holds

$$\int_{\mathcal{X}^n} f(\mathbf{x}) P_\xi^n(\mathrm{d}\mathbf{x}) = \int_{\mathcal{X}^n} \int_{\mathcal{P}} f(\mathbf{x}) q(\mathrm{d}x_1) \cdots q(\mathrm{d}x_n) \pi_n(\mathrm{d}q).$$

Indeed, if $f = \mathbb{1}_A$, then

$$P_\xi^n(A) = \int_{\mathcal{P}} q^n(A) \pi_n(\mathrm{d}q) = \int_{\mathcal{P}} \int_{\mathcal{X}^n} f(\mathbf{x}) q(\mathrm{d}x_1) \cdots q(\mathrm{d}x_n) \pi_n(\mathrm{d}q).$$

By linearity, the same holds for simple functions and then extends to nonnegative measurable functions by using the monotone convergence theorem and Fubini-Tonelli's theorem.

Thus, if $A = A_1 \times \cdots \times A_n$, then

$$P_n(A) = \frac{1}{Z_n} \int_{\mathcal{P}} \Big(\prod_{j=1}^n \int_{A_j} w_n(x) q(\mathrm{d}x) \Big) \pi_n(\mathrm{d}q), \tag{2.3}$$

and now it is easy to see that

$$Z_n = \int_{\mathcal{P}} \langle w_n, q \rangle^n \pi_n(\mathrm{d}q). \tag{2.4}$$

An alternative representation of P_n. It turns out that it is convenient to represent the measure P_n slightly differently as follows. For every $q \in \mathcal{P}$, we define a probability measure $r_{q,n}$ on $\mathcal{B}(\mathcal{X})$:

$$r_{q,n}(A) := \frac{\int_A w_n(x) q(\mathrm{d}x)}{\langle w_n, q \rangle}, \quad A \in \mathcal{B}(\mathcal{X}). \tag{2.5}$$

The mapping $q \mapsto \langle w_n, q \rangle$ is continuous, and it can also be shown that for any fixed $A \in \mathcal{B}(\mathcal{X})$ the mapping $q \mapsto \int_A w_n(x) q(\mathrm{d}x)$ is measurable. Indeed, $q \mapsto q(A)$ is measurable for all $A \in \mathcal{B}(\mathcal{X})$, hence by linearity $q \mapsto \int_{\mathcal{X}} f(x) q(\mathrm{d}x)$ is measurable for all measurable simple function f; for a generic nonnegative measurable function f, the result follows by taking the usual limit argument $f_n \uparrow f$ where $\{f_n\}_n$ are simple functions. And so for any A, the mapping $q \mapsto r_{q,n}(A)$ is measurable as well. By $\pi - \lambda$ argument, for any $A \in \mathcal{B}(\mathcal{X}^n)$, $q \mapsto r_{q,n}^n(A)$ is measurable, where $r_{q,n}^n$ stands for n-fold product measure.

Given a probability measure π_n on $\mathcal{B}(\mathcal{P})$, we define another probability measure $\bar{\pi}_n$ on $\mathcal{B}(\mathcal{P})$ as follows

$$\bar{\pi}_n(E) := \frac{1}{Z_n} \int_E \langle w_n, q \rangle^n \pi_n(\mathrm{d}q), \quad E \in \mathcal{B}(\mathcal{P}). \tag{2.6}$$

Here Z_n is the normalizing constant, thus Z_n is as in (2.4). Now, the measure P_n can be alternatively defined as follows

$$P_n(A) = \int_{\mathcal{P}} r_{q,n}^n(A) \bar{\pi}_n(\mathrm{d}q), \quad A \in \mathcal{B}(\mathcal{X}^n). \tag{2.7}$$

To see that the equality (2.7) holds, observe that the right hand side of (2.7) defines a probability measure that for any measurable cylinder $A = A_1 \times \cdots \times A_n$ reads

$$\int_{\mathcal{P}} r_{q,n}^n(A) \bar{\pi}_n(\mathrm{d}q) = \int_{\mathcal{P}} \prod_{i=1}^n r_{q,n}(A_i) \bar{\pi}_n(\mathrm{d}q) = \frac{1}{Z_n} \int_{\mathcal{P}} \Big(\prod_{i=1}^n \int_{A_i} w_n(x) q(\mathrm{d}x) \Big) \pi_n(\mathrm{d}q) = P_n(A),$$

where the last equality holds by (2.3). Therefore these measures coincide on cylinders, hence also on $\mathcal{B}(\mathcal{X}^n)$.

We can go one step further, and consider the mapping

$$r_n : \mathcal{P} \mapsto \mathcal{P}, \quad r_n(q) := r_{q,n}. \tag{2.8}$$

For every n, the map r_n is continuous: let f be a bounded continuous function on $\mathcal{X}$. Note that w_n and $f \cdot w_n$ and bounded and continuous for every n. Whence,

$$\int_{\mathcal{X}} f(x) r_{q_m,n}(\mathrm{d}x) = \frac{1}{\langle w_n, q_m \rangle} \int_{\mathcal{X}} f(x) w_n(x) q_m(\mathrm{d}x) \stackrel{m\to\infty}{\longrightarrow} \frac{1}{\langle w_n, q \rangle} \int_{\mathcal{X}} f(x) w_n(x) q(\mathrm{d}x) = \int_{\mathcal{X}} f(x) r_{q,n}(\mathrm{d}x).$$

Thus $r_n(q_m) = r_{q_m,n} \Rightarrow r_{q,n} = r_n(q)$, whence r_n is measurable. Now define the pushforward measure ν_n on $\mathcal{B}(\mathcal{P})$ as follows $\nu_n(E) := \bar{\pi}_n\big(r_n^{-1}(E)\big)$. Thus by change of variable formula

$$P_n(A) = \int_{\mathcal{P}} r^n_{q,n}(A) \bar{\pi}_n(\mathrm{d}q) = \int_{\mathcal{P}} q^n(A) \nu_n(\mathrm{d}q), \quad A \in \mathcal{B}(\mathcal{X}^n). \tag{2.9}$$

The measure Q_n. Recall the mapping g defined in (1.5) and the measure Q_n defined in (1.6). The mapping g is many-to-one, because all permutation of a vector $\mathbf{x}$ have the same $g(\mathbf{x})$. Observe that for any function f on $\mathcal{X}^n$, it holds: $n^{-1}\sum_{i=1}^n f(x_i) = \langle f, g(\mathbf{x})\rangle$. This observation helps us to see that the mapping g is continuous. Let $\mathbf{x}^m \to \mathbf{x}$ be a convergent sequence in $\mathcal{X}^n$. Since the convergence in $\mathcal{X}^n$ is equivalent to pointwise convergence, it follows that as $m \to \infty$, for any continuous and bounded function on $\mathcal{X}^n$, i.e for any $f \in C_b(\mathcal{X}^n)$, it holds

$$\langle f, g(\mathbf{x}^m)\rangle = n^{-1}\sum_{i=1}^n f(x_i^m) \to n^{-1}\sum_{i=1}^n f(x_i) = \langle f, g(\mathbf{x})\rangle.$$

So the convergence $\mathbf{x}^m \to \mathbf{x}$ implies $g(\mathbf{x}^m) \Rightarrow g(\mathbf{x})$, hence g is continuous and measurable. The advantage of Q_n over P_n is that, for every n, the measure Q_n is defined on the same domain $\mathcal{B}(\mathcal{P})$, and so one can study the convergence on Q_n in the usual sense of weak convergence of probability measures. When $\mathcal{X}$ is finite, then the measure Q_n can be constructed explicitly, see Lember and Watkins (2022).

3. Limit process and limit measure Q^*

3.1. *The limit process.* We now turn to the asymptotics of P_n as n grows. Recall that we aim to show the existence of a limit process $X_1, X_2, \ldots$ so that (1.4) holds, where $(X_{1,n}, \ldots, X_{n,n}) \sim P_n$. Observe that (1.4) is equivalent to the following: for any $m \in \mathbb{N}$,

$$(X_{1,n}, \ldots, X_{m,n}) \Rightarrow (X_1, \ldots, X_m). \tag{3.1}$$

Indeed, from (3.1), it follows that (1.4) holds when $t_1 < t_2 < \ldots < t_m$ and the weak convergence of random vectors implies that of the permutations. According to (2.7), for every $A_i \in \mathcal{B}(\mathcal{X})$, $i = 1, \ldots, m$, it holds

$$\mathbf{P}(X_{1,n} \in A_1, \ldots, X_{m,n} \in A_m) = \int_{\mathcal{P}} \prod_{i=1}^m r_{q,n}(A_i) \bar{\pi}_n(\mathrm{d}q) =: P_n(A_1 \times \cdots \times A_m).$$

By the canonical representation, the existence of a stochastic process is equivalent to the existence of a probability measure P^* on $(\mathcal{X}^\infty, \Sigma)$, where Σ is the product σ-algebra. The measure P^* can be considered as the distribution of X. Let $\mathcal{C}(A_1 \times \cdots \times A_m) := \{(x_i) \in \mathcal{X}^\infty : x_1 \in A_1, \ldots, x_m \in A_m\}$ be any measurable cylinder, where $A_i \in \mathcal{B}(\mathcal{X})$, $i = 1, \ldots, m$. With slight abuse of notation, we shall denote by $P^*(A_1 \times \cdots \times A_m)$ the measure of the cylinder $\mathcal{C}(A_1 \times \cdots \times A_m)$. If P^* is the distribution of X, then $P^*(A_1 \times \cdots \times A_m) = \mathbf{P}(X_1 \in A_1, \ldots, X_m \in A_m)$. Since cylinders are a convergence-determining class, (Billingsley, 1999, Theorem 2.8), the convergence (3.1) holds if for every m, and every measurable and P^*-continuous cylinder $A = \{(x_i) \in \mathcal{X}^\infty : x_1 \in A_1, \ldots, x_m \in A_m\}$ it holds

$$P_n(A_1 \times \cdots \times A_m) \to P^*(A_1 \times \cdots \times A_m). \tag{3.2}$$

Recall that A is P^* continuous when $P^*(\partial A) = 0$, where ∂A stands for the boundary of A. To summarize, for showing (1.4), it suffices to show the existence of a probability measure P^* on $(\mathcal{X}^\infty, \Sigma)$ such that for all P^*-continuous cylinders (3.2) holds.

In the following theorem $\{r_{q,n}\}_n$ are probability measures on $\mathcal{X}$ which do not necessarily coincide with the ones defined by (2.5) (with w_n as in (1.7)). We will see that for that particular choice of measures $\{r_{q,n}\}_n$, by Corollary 3.2, the hypotheses in Theorem 3.1 concerning uniform convergence to r_q and continuity of the map $q \to r_q$ are always satisfied.

Theorem 3.1. *Suppose there exists a probability measure $\bar\pi$ on $\mathcal{P}$ such that $\bar\pi_n \Rightarrow \bar\pi$. For every $q \in \mathcal{P}$ and $n \in \mathbb{N}$, let $r_{q,n}$, $r_q \in \mathcal{P}$ be such that $\sup_{q\in\mathcal{P}} |r_{q,n}(A) - r_q(A)| \to 0$ for all $A \in \mathcal{B}(\mathcal{X})$. Assume also that $q \mapsto r_q$ is continuous. Then there exists an infinitely exchangeable process X so that for every $m \in \mathbb{N}$, the convergence (3.1) holds. Moreover, the limit process is such that for every $m \in \mathbb{N}$ and $A_1, \dots, A_m \in \mathcal{B}(\mathcal{X})$,*

$$P^*\big(A_1 \times \cdots \times A_m\big) := \int_{\mathcal{P}} r_q^m\big(A_1 \times \cdots \times A_m\big)\bar\pi(\mathrm{d}q). \tag{3.3}$$

Proof: Let for every m and distinct integers $t_1, \dots, t_m \in \mathbb{N}$, $\mu_{t_1,\dots,t_m}$ be a probability measure on $\mathcal{B}(\mathcal{X}^m)$ defined as follows:

$$\mu_{t_1,\dots,t_m}(A) := \int_{\mathcal{P}} r_q^m(A)\bar\pi(\mathrm{d}q), \quad A \in \mathcal{B}(\mathcal{X}^m).$$

The definition is correct, because by assumption $q \mapsto r_q$ is measurable, and so for every m and $A \in \mathcal{B}(\mathcal{X}^m)$, the mapping $q \mapsto r_q^m(A)$ (product measure) is measurable as well. Note that this definition depends on m but is independent of the choice of $t_1, \dots, t_m$. Clearly the family $\{\mu_{t_1,\dots,t_m}\}$ fulfills the consistency conditions, and so by Kolmogorov existence theorem there exists a measure P^* on $(\mathcal{X}^\infty, \Sigma)$ such that for every distinct integers $t_1, \dots, t_m$ and every $A \in \mathcal{B}(\mathcal{X}^m)$, it holds

$$P^*\big(\{(x_i) \in \mathcal{X}^\infty : (x_{t_1}, \dots, x_{t_m}) \in A\}\big) = \mu_{t_1,\dots,t_m}(A).$$

In particular

$$P^*(A_1 \times \cdots \times A_m) = \mu_{1,\dots,n}(A_1 \times \cdots \times A_m) = \int_{\mathcal{P}} \prod_{i=1}^m r_q(A_i)\bar\pi(\mathrm{d}q).$$

Thus P^* is the distribution of an infinitely exchangeable process, and the theorem is proven, when we show that (3.2) holds for all P^*-continuous cylinders. To show (3.2), we use Skorohod representation theorem (Billingsley, 1999, Theorem 6.7) according to which there are $\mathcal{P}$-valued random variables W_n and W such that W_n has distribution $\bar\pi_n$, W has distribution $\bar\pi$ and $W_n \to W$ a.s.. The theorem applies on separable metric space, but $\mathcal{P}$ equipped with Prokhorov metric is separable. Fix a P^*-continuous cylinder $A = \{(x_i) \in \mathcal{X}^\infty : x_1 \in A_1, \dots, x_m \in A_m\}$ and let us denote

$$f_n(q) := \prod_{i=1}^m r_{q,n}(A_i) = r_{q,n}^m(A_1 \times \cdots \times A_m), \quad f(q) := \prod_{i=1}^m r_q(A_i) = r_q^m(A_1 \times \cdots \times A_m).$$

If $m = 1$, then by assumption, it immediately follows that $\sup_q |f_n(q) - f(q)| \to 0$. Since the functions are bounded, the uniform convergence also holds when $m > 1$. Indeed, if $f_n \to f$ and $g_n \to g$ uniformly and all functions are bounded by 1, then

$$|f_n g_n - fg| = |f_n g_n - f_n g + f_n g - fg| \le |f_n(g_n - g)| + |g(f_n - f)| \le |f_n - f| + |g_n - g|.$$

Let $E_{\text{cont}} \subset \mathcal{P}$ be the set of continuity points of f. We shall show that $\bar\pi(E_{\text{cont}}) = 1$. Since

$$\partial A = \{(x_i) \in \mathcal{X}^\infty : (x_1, \dots, x_m) \in \partial(A_1 \times \cdots \times A_m)\},$$

where $\partial(A_1 \times \cdots \times A_m)$ is the boundary in $\mathcal{X}^m$, we have (since A is P^*-continuous)

$$P^*(\partial A) = \int_{\mathcal{P}} r_q^m\big(\partial(A_1 \times \cdots \times A_m)\big)\bar\pi(\mathrm{d}q) = 0.$$

The integral of non-negative function is zero only if the function is $\bar{\pi}$-a.s. equal to 0 and so we have $\bar{\pi}(F) = 1$, where

$$F := \{q : r_q^m\big(\partial(A_1 \times \cdots \times A_m)\big) = 0\}.$$

We now show that $F \subset E_{\text{cont}}$. Indeed, if $q \in F$ and $q_n \Rightarrow q$, then by the continuity assumption $r_{q_n} \Rightarrow r_q$ and thus also $r_{q_n}^m \Rightarrow r_q^m$ (because weak convergence of marginal measures implies weak convergence of the product measure). Since $r_q^m\big(\partial(A_1 \times \cdots \times A_m)\big) = 0$, it follows that

$$r_{q,n}^m(A_1 \times \cdots \times A_m) \to r_q^m(A_1 \times \cdots \times A_m).$$

Thus $f(q_n) \to f(q)$ and so $q \in E_{\text{cont}}$. Since $\bar{\pi}(E_{\text{cont}}) = 1$, from $W_n \to W$ a.s. it follows $f(W_n) \to f(W)$ a.s. From the uniform convergence, it follows that $|f_n(W_n) - f(W_n)| \to 0$. These two facts together imply $f_n(W_n) \to f(W)$, a.s. Finally, since the functions f_n are all bounded by 1, by the bounded convergence theorem it follows $Ef_n(W_n) \to Ef(W)$. Since

$$P_n(A_1 \times \cdots \times A_m) = \int_{\mathcal{P}} r_{q,n}^m(A_1 \times \cdots \times A_m)\bar{\pi}_n(\mathrm{d}q) = \int_{\mathcal{P}} f_n(q)\bar{\pi}_n(\mathrm{d}q) = Ef_n(W_n)$$
$$P^*(A_1 \times \cdots \times A_m) = \int_{\mathcal{P}} r_q^m(A_1 \times \cdots \times A_m)\bar{\pi}(\mathrm{d}q) = \int_{\mathcal{P}} f(q)\bar{\pi}(\mathrm{d}q) = Ef(W),$$

we have (3.2). □

Corollary 3.2. *Let $r_{q,n}$ and w_n be as defined in (2.5) and (1.7) respectively. If $\lambda > 0$, define $r_q = q$, while if $\lambda = 0$, let r_q be the measure proportional to $w\mathrm{d}q$, where $w = w_n$ (in this case w_n does not depend on n). Then*

1. $\sup_{q \in \mathcal{P}} |r_{q,n}(A) - r_q(A)| \to 0$ *for all $A \in \mathcal{B}(\mathcal{X})$;*
2. *$q \mapsto r_q$ is continuous;*
3. *if $\bar{\pi}_n \Rightarrow \bar{\pi}$ (where $\bar{\pi}_n$ is defined in (2.6)) then P_n converges (in the sense of (3.1)) to the measure P^* defined in (3.3).*

Proof: (1) Recall that $q \mapsto r_{q,n}$ is continuous as explained after (2.8) and that $w_n(x) = \exp(-\phi(x)/n^\lambda)$. Let $w(x) := 1$ (for all x) if $\lambda > 0$ and define $w(x) := \exp(-\phi(x)) = w_n(x)$ (for all x) if $\lambda = 0$.
Note that $\sup_{x \in \mathcal{X}} |w_n(x) - w(x)| = 0$ if $\lambda = 0$. If $\lambda > 0$, then $\sup_{x \in \mathcal{X}} |w_n(x) - w(x)| = 1 - \exp(-\sup_x \phi(x)/n^\lambda)$, which goes to 0 as n tends to infinity, since ϕ is by hypothesis bounded. It follows that

$$\sup_q |\langle w_n, q\rangle - \langle w, q\rangle| \leq \sup_q \langle |w_n - w|, q\rangle \to 0.$$

Clearly the same argument holds when integrating over any set A. We now observe that $\inf_{x \in \mathcal{X}} w_n(x) = \exp(-\sup_x \phi(x)/n^\lambda) > 0$, when $\lambda > 0$ and it is equal to 1 when $\lambda = 0$. This implies that, for all $A \in \mathcal{B}(\mathcal{X})$,

$$\sup_{q \in \mathcal{P}} |r_{q,n}(A) - r_q(A)| = \sup_{q \in \mathcal{P}} \left| \frac{\langle \mathbb{1}_A w_n, q\rangle}{\langle w_n, q\rangle} - \frac{\langle \mathbb{1}_A w, q\rangle}{\langle w, q\rangle} \right| \to 0.$$

(2) Although the continuity of $q \mapsto r_q$ follows by the same argument as the continuity of $q \mapsto r_{q,n}$, we shall now show that it can be directly deduced from the continuity of $q \mapsto r_{q,n}$ and the uniform convergence stated in 1. Let f be a bounded, nonnegative, measurable function on $\mathcal{X}$ and consider a sequence $\{q_m\}_m$ such that $q_m \Rightarrow q$. By using the Uniform Bounded Convergence Theorem (see for instance Bertacchi and Zucca, 2003, Theorem 2.3), as $n \to \infty$,

$$\sup_{q \in \mathcal{P}} \left| \int_{\mathcal{X}} f(x) r_{q,n}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_q(\mathrm{d}x) \right| = \sup_{q \in \mathcal{P}} \left| \int_0^{+\infty} r_{q,n}(f \geq t)dt - \int_0^{+\infty} r_q(f \geq t)dt \right| \to 0.$$

Suppose now that, in addition, f is continuous. Take $\varepsilon > 0$ and $n_0 = n_0(\varepsilon)$ such that for all $n \geq n_0$ we have

$$\sup_{q\in\mathcal{P}} \Big| \int_{\mathcal{X}} f(x) r_{q,n}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_q(\mathrm{d}x) \Big| < \varepsilon/3.$$

Since $q \mapsto r_{q,n}$ is continuous for every $n \in \mathbb{N}$, take $m_0 = m_0(\varepsilon, n_0)$ such that

$$\Big| \int_{\mathcal{X}} f(x) r_{q_m,n_0}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_{q,n_0}(\mathrm{d}x) \Big| < \varepsilon/3$$

for all $m \geq m_0$. Clearly, for all $m \geq m_0$

$$\begin{aligned}\Big| \int_{\mathcal{X}} f(x) r_{q_m}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_q(\mathrm{d}x) \Big| \leq& \Big| \int_{\mathcal{X}} f(x) r_{q_m}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_{q_m,n_0}(\mathrm{d}x) \Big| \\ &+ \Big| \int_{\mathcal{X}} f(x) r_{q_m,n_0}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_{q,n_0}(\mathrm{d}x) \Big| \\ &+ \Big| \int_{\mathcal{X}} f(x) r_{q,n_0}(\mathrm{d}x) - \int_{\mathcal{X}} f(x) r_q(\mathrm{d}x) \Big| < \varepsilon.\end{aligned}$$

(3) If $\bar{\pi}_n \Rightarrow \bar{\pi}$ then all the assumptions of Theorem 3.1 are satisfied and the claim follows.

□

Remark 3.3. The proof of Corollary 3.2 shows that the assumptions of Theorem 3.1 on $r_{q,n}$ and r_q are satisfied also with more general weight functions than those considered in (1.7). Indeed

- if the weight functions are such that $q \mapsto r_{q,n}$ is continuous for every $n \in \mathbb{N}$ and $\sup_{q\in\mathcal{P}} |r_{q,n}(A) - r_q(A)| \to 0$ for all $A \in \mathcal{B}(\mathcal{X})$, then $q \mapsto r_q$ is continuous;
- in particular, if w_n and w are measurable functions so that $\inf_{x\in\mathcal{X}} w(x) > 0$ and $\sup_{x\in\mathcal{X}} |w_n(x) - w(x)| \to 0$, then $q \mapsto r_{q,n}$ is continuous for every $n \in \mathbb{N}$ and $\sup_{q\in\mathcal{P}} |r_{q,n}(A) - r_q(A)| \to 0$ for all $A \in \mathcal{B}(\mathcal{X})$ with r_q being the probability measure proportional to $w\mathrm{d}q$.

3.2. *The weak convergence of* Q_n. The goal is to show that under the same assumptions as in Theorem 3.1 with the additional requirement that $\mathcal{X}$ is compact, the measures Q_n converge to a measure Q. Recall the function g in (1.5) that maps every sequence $\mathbf{x} = (x_1, \dots, x_n)$ to its empirical measure. To stress the dependence of n, it this section, we shall denote the function g as g_n.

Theorem 3.4. *Suppose $\mathcal{X}$ is compact and the assumptions of Theorem 3.1 hold. Let $Q = \bar{\pi} r^{-1}$, where $r : \mathcal{P} \mapsto \mathcal{P}$ is defined from the r_q in Theorem 3.1, as $r(q) = r_q$ for all $q \in \mathcal{P}$. Then $Q_n \Rightarrow Q$.*

Proof: By the Portmanteau theorem (see for instance Billingsley, 1999, Theorem 2.1), it suffices to show that for every open set $E \in \mathcal{B}(\mathcal{P})$ we have $\liminf_n Q_n(E) \geq Q(E)$. Recall that, according to (2.9),

$$Q_n(E) = \int_{\mathcal{P}} r_{q,n}^n \big(g_n^{-1}(E)\big) \bar{\pi}_n(\mathrm{d}q) = \int_{\mathcal{P}} q^n \big(g_n^{-1}(E)\big) \nu_n(\mathrm{d}q), \quad E \in \mathcal{B}(\mathcal{P}).$$

Let E be an open set. We first show that $\liminf_n \nu_n(E) \geq Q(E)$. For that, we use Skorohod representation again, so let $Z_n \sim \bar{\pi}_n$, $Z \sim \bar{\pi}$ be $\mathcal{P}$-valued random variables so that $Z_n \to Z$, a.s.. Recall that we use parallel notation $r_n(q) :=: r_{q,n}$ and $r(q) :=: r_q$, where $r_n, r : \mathcal{P} \to \mathcal{P}$. We now argue that the a.s. convergence $Z_n \to Z$ (with respect to Prokhorov metric) entails $r_n(Z_n) \to r(Z)$, a.s.. Since the convergence with respect to Prokhorov metric is equivalent to to the weak convergence of measures, it suffices to show that $q_n \Rightarrow q$ implies $r_n(q_n) \Rightarrow r(q)$. To see that take A to be a r_q-continuous set. Then

$$|r_{q_n,n}(A) - r_q(A)| \leq |r_{q_n,n}(A) - r_{q_n}(A)| + |r_{q_n}(A) - r_q(A)|.$$

By assumption, $|r_{q_n,n}(A) - r_{q_n}(A)| \leq \sup_q |r_{q,n}(A) - r_q(A)| \to 0$, and since $q \mapsto r(q)$ is continuous, it follows that $|r_{q_n}(A) - r_q(A)| \to 0$. Hence $r_n(q_n) \Rightarrow r(q)$, and so $r_n(Z_n) \to r(Z)$. Since E is open, it follows that

$$P\big(\{r(Z) \in E\} \setminus \liminf_n \{r_n(Z_n) \in E\}\big) = 0$$

and so the following holds for all open sets E

$$\begin{aligned} Q(E) = \bar{\pi}\big(r^{-1}(E)\big) &= P(r(Z) \in E) \leq P\big(\liminf_n \{r_n(Z_n) \in E\}\big) \\ &\leq \liminf_n P(r_n(Z_n) \in E) = \liminf_n \bar{\pi}_n\big(r_n^{-1}(E)\big) = \liminf_n \nu_n(E). \end{aligned} \tag{3.4}$$

Denote by $m(\delta)$ the δ-covering number (i.e. the minimal number of δ-balls needed to cover $\mathcal{P}$), $E_\delta = \{p \in \mathcal{P} : d(p, E) \leq \delta\}$ the closed δ-blowup of E and recall the definition of relative entropy

$$D(p\|q) := \begin{cases} \int_{\mathcal{X}} \ln\big(\frac{dp}{\mathrm{d}q}\big) dp & \text{if } p \ll q \\ +\infty & \text{else.} \end{cases} \tag{3.5}$$

When $\mathcal{X}$ is compact, then also $\mathcal{P}$ is compact, so $m(\delta) < \infty$. For us it is important that $m(\delta)$ is independent of q.
Since $\mathcal{X}$ is compact, for every $q \in \mathcal{P}$ the following inequality holds (see Dembo and Zeitouni, 2010, Example 6.2.19):

$$q^n\big(g_n^{-1}(E)\big) \leq \inf_{\delta>0} \Big(m(\delta) \exp[-\inf_{p \in E_\delta} D(p\|q) \cdot n]\Big). \tag{3.6}$$

Let E_δ^c be closed δ-blowup of E^c. Then, for any $\delta > 0$, define

$$F_\delta := \big(E_{2\delta}^c\big)^c.$$

Clearly F_δ is an open set inside E, and $\cup_{\delta>0} F_\delta = E$. We now argue that for any $\delta > 0$,

$$\inf_{p \in E_\delta^c, q \in F_\delta} D(p\|q) > 0. \tag{3.7}$$

If not, there would be a sequence $\{q_n\}_n$ in F_δ and $\{p_n\}_n$ in E_δ^c so that $D(p_n\|q_n) \to 0$. From Pinsker's inequality, it follows that $d(p_n, q_n) \to 0$, where d stands for Prokhorov metric. But since $p_n \in E_\delta^c$ and $q_n \in \big(E_{2\delta}^c\big)^c$, it must be that $d(p_n, q_n) > \delta$ for every n. Hence (3.7) holds. From (3.6), we obtain that for every $\delta > 0$,there exists $\varepsilon(\delta)$ such that

$$\sup_{q \in F_\delta} q^n\big(g_n^{-1}(E^c)\big) \leq m(\delta) \exp[-\varepsilon(\delta) \cdot n], \quad \inf_{q \in F_\delta} q^n\big(g_n^{-1}(E)\big) \geq 1 - m(\delta) \exp[-\varepsilon(\delta) \cdot n]. \tag{3.8}$$

Therefore,

$$Q_n(E) \geq \int_{F_\delta} q^n(g_n^{-1}(E)) \nu_n(\mathrm{d}q) \geq \big(1 - m(\delta) \exp[-\varepsilon(\delta) \cdot n]\big) \nu_n(F_\delta).$$

From equation (3.4), since F_δ is open we deduce that

$$\liminf_n Q_n(E) \geq \liminf_n \nu_n(F_\delta) \geq Q(F_\delta) \uparrow Q(E), \quad \text{as } \delta \downarrow 0$$

where the last limit follows from the continuity of a measure. □

Remark 3.5. The additional assumption of compactness is disappointing. In the proof above, it was needed for the Sanov type of inequality (3.6), which, in turn, was needed for the uniform convergence in (3.8). The convergence (3.8) is, in a sense, exponentially fast, but for our proof the speed is not important, it just suffices to have:

$$\inf_{q \in F_\delta} q^n(g_n^{-1}(E)) \to 1. \tag{3.9}$$

Recall that for every open E, and for every $q \in E$, by SLLN $q^n(g_n^{-1}(E)) \to 1$. The convergence (3.9) states that on the set F_δ this convergence is uniform and then our proof applies.

Remark 3.6. In Lember and Watkins (2022), it was shown that, when $\mathcal{X}$ is finite, for any continuous and bounded $f : \mathcal{P} \to \mathbb{R}$, it holds: $\int_{\mathcal{P}} f dQ_n \to \int_{\mathcal{P}} f dQ$. In our notation

$$\int_{\mathcal{P}} f(q) Q_n(\mathrm{d}q) = \int_{\mathcal{P}} \int_{\mathcal{X}^n} f(g_n(\mathbf{x})) q^n(\mathrm{d}\mathbf{x}) \nu_n(dq).$$

By SLLN,

$$f_n(q) := \int_{\mathcal{X}^n} f(g_n(\mathbf{x})) q^n(\mathrm{d}\mathbf{x}) \to f(q),$$

and if this convergence were uniform, i.e.

$$\sup_q |f_n(q) - f(q)| \to 0, \tag{3.10}$$

then from $\nu_n \Rightarrow Q$, by using the Skorohod representation, it would follow that $\int_{\mathcal{P}} f_n(q)\nu_n(\mathrm{d}q) \to \int_{\mathcal{P}} f(q)Q(\mathrm{d}q)$. In Lember and Watkins (2022) the equation (3.10) for finite $\mathcal{X}$ was obtained with Bernstein polynomials.

4. Arbitrary prior π

In this section, we consider the case when the prior π is arbitrary and independent of n. In what follows, we take w_n as defined in (1.7). By Theorem 3.1 and Corollary 3.2, to ensure the existence of the limit process, it suffices to show $\bar{\pi}_n \Rightarrow \bar{\pi}$.

Since in the integral (2.6) defining $\bar{\pi}_n$ we have the function $\langle w_n, q\rangle^n$, we start with the following observation.

Proposition 4.1. *If $m \to \infty$, and ϕ is nonnegative and integrable with respect to q (but not necessarily bounded), then*

$$\langle \exp[-\frac{\phi}{m}], q\rangle^m \to \exp[-\langle \phi, q\rangle]. \tag{4.1}$$

Moreover, when ϕ is bounded, then the convergence is uniform over q.

Proof: Let us consider an i.i.d. sequence $\{X_i\}_{i\in\mathbb{N}}$ of random variables with law q. Clearly, by using the Law of Large Numbers and the Bounded Convergence Theorem, we have

$$\begin{aligned}\Big(\int_{\mathcal{X}} \exp\Big(-\frac{\phi(x)}{m}\Big) q(\mathrm{d}x)\Big)^m &= \mathbb{E}\Big[\prod_{i=1}^m \exp\Big(-\frac{\phi(X_i)}{m}\Big)\Big] \\ &= \mathbb{E}\Big[\exp\Big(-\sum_{i=1}^m \frac{\phi(X_i)}{m}\Big)\Big] \to \exp\big(\mathbb{E}[-\phi(X_1)]\big)\end{aligned}$$

since $\exp(-\sum_{i=1}^m \frac{\phi(X_i)}{m}) \to \exp(\mathbb{E}[-\phi(X_1)])$ a.s. as $m \to \infty$.

Suppose now that ϕ is bounded, say $|\phi(x)| \le K$ for all $x \in \mathcal{X}$; it is enough to prove that $\mathbb{E}[|Y_n - m_q|] \to 0$ as $n \to +\infty$ uniformly with respect to q, where $Y_n := \sum_{i=1}^n \phi(X_i)/n$ and $m_q := \mathbb{E}[\phi(X_1)]$. For every $\varepsilon > 0$, given any q we have

$$\begin{aligned}\mathbb{E}[|Y_n - m_q|] &= \mathbb{E}[|Y_n - m_q|\mathbb{1}_{\{|Y_n - m_q| \le \varepsilon/2\}}] + \mathbb{E}[|Y_n - m_q|\mathbb{1}_{\{|Y_n - m_q| > \varepsilon/2\}}] \\ &\le \frac{\varepsilon}{2} + 2KP\big(|Y_n - m_q| > \varepsilon/2\big) \le \frac{\varepsilon}{2} + \frac{8K^3}{n(\varepsilon/2)^2} \le \varepsilon\end{aligned}$$

if $n \ge \frac{8K^3}{(\varepsilon/2)^3}$ (and this does not depend on q). □

4.1. *The case* $\lambda \geq 1$. Let us begin with the case $\lambda \geq 1$. We establish the convergence of the measure $\bar{\pi}_n$ which was defined in (2.6). The next theorem states that when $\lambda > 1$, then the influence of fitness vanishes, and the limit process X equals the breeding process ξ. When $\lambda = 1$, then the limit process is another infinitely exchangeable process whose prior measure differs from the breeding one π, and depends on ϕ as well as on π.

Theorem 4.2. *Let the fitness function be as in* (1.7) *where* ϕ *is non-negative, measurable and bounded. Suppose* $\pi_n = \pi$*; and let* $\lambda \geq 1$*. Then the following convergences hold:*

(1) *If* $\lambda = 1$*, then* $\bar{\pi}_n \Rightarrow \bar{\pi}$*, and* $P_n \to P^*$ *in the sense of* (3.1)*, where*

$$\bar{\pi}(E) := \frac{1}{Z}\int_E \exp[-\langle \phi, q\rangle]\pi(\mathrm{d}q), \quad \text{where} \quad Z := \int_{\mathcal{P}} \exp[-\langle \phi, q\rangle]\pi(\mathrm{d}q),$$

$$P^*(A_1 \times \cdots \times A_m) = \int_{\mathcal{P}} \prod_{i=1}^m q(A_i)\bar{\pi}(\mathrm{d}q), \quad \forall m \in \mathbb{N}, A_i \in \mathcal{B}(\mathcal{X}).$$

If, in addition, $\mathcal{X}$ *is compact, then* $Q_n \Rightarrow \bar{\pi}$*.*

(2) *If* $\lambda > 1$*, then* $\bar{\pi}_n \Rightarrow \pi$ *and* $P_n \to P_\xi$ *in the sense of* (3.1)*. If, in addition,* $\mathcal{X}$ *is compact, then* $Q_n \Rightarrow \pi$*.*

Proof: Before explicitly dealing with the two cases, we note that by Corollary 3.2, we only need to establish the convergence of $\bar{\pi}_n$. Since $r_q = q$, from Theorem 3.1 we will get the convergence of P_n to the limiting process.

(1) Since for any q and any n, it holds $\langle \exp[-\frac{\phi}{n}], q\rangle^n \leq 1$, we obtain from (4.1) and the Bounded Convergence Theorem that, for any $E \in \mathcal{B}(\mathcal{P})$,

$$Z_n\bar{\pi}_n(E) = \int_E \langle \exp[-\frac{\phi}{n}], q\rangle^n \pi(\mathrm{d}q) \to \int_E \exp[-\langle \phi, q\rangle]\pi(\mathrm{d}q). \tag{4.2}$$

From (4.2), it follows that $\bar{\pi}_n(E) \to \bar{\pi}(E)$, meaning that $\bar{\pi}_n \Rightarrow \bar{\pi}$. Since the assumptions of Theorem 3.1 are fulfilled with $w(x) = 1$, then for compact $\mathcal{X}$ Theorem 3.4 implies $Q_n \Rightarrow \bar{\pi}$.

(2) When $\lambda > 1$, then by Proposition 4.1, eventually as $n \to \infty$

$$1 \leftarrow \big(2\exp(-\langle \phi, q\rangle)\big)^{1/n^{\lambda-1}} \geq \langle \exp[-\frac{\phi}{n^\lambda}], q\rangle^n \geq \Big(\frac{\exp(-\langle \phi, q\rangle)}{2}\Big)^{1/n^{\lambda-1}} \to 1$$

and by dominated convergence, again, for any measurable E

$$\int_E \langle \exp[-\frac{\phi}{n}], q\rangle^n \pi(\mathrm{d}q) \to \pi(E). \tag{4.3}$$

Therefore $\bar{\pi}_n \Rightarrow \pi$. The convergence $Q_n \Rightarrow \pi$ is a consequence of Theorem 3.4.

□

Observe that for the weak convergence of $\bar{\pi}_n$, the boundedness of ϕ is not needed. However, it is needed for the uniform convergence of $w_n \to w$, hence for existence of the limit process (Theorem 3.1) and for the weak convergence of Q_n (Theorem 3.4). Theorem 4.2 is a direct generalization of Theorem 5.1 (2) and (3) in Lember and Watkins (2022), and no additional assumptions are imposed.

4.2. *Case* $\lambda \in [0, 1)$.

Preliminaries: densities and powers. Let π be a finite measure (not necessarily a probability measure) on $\mathcal{B}(\mathcal{P})$. Let $\mathcal{S}$ be the support of π. For a measurable function $f : \mathcal{P} \to \mathbb{R}$,

$$\|f\|_\infty := \mathrm{esssup}(f) := \inf\{c : |f| \leq c \quad \pi - \text{a.e.}\}.$$

If f is continuous then $\|f\|_\infty = \sup_{q\in\mathcal{S}} |f(q)|$. Also recall that for any $0 < m < \infty$

$$\|f\|_m := \Big(\int_{\mathcal{P}} |f(q)|^m \pi(\mathrm{d}q) \Big)^{\frac{1}{m}}.$$

If f is essentially bounded, m grows and π is a probability measure, then $\|f\|_m \nearrow \|f\|_\infty < \infty$. Then it follows that $\|f\|_m \to \|f\|_\infty$ also when π is a finite (but not necessarily a probability) measure.

For any sequence measurable, essentially bounded functions $f_n : \mathcal{P} \to \mathbb{R}^+$, and $m_n \uparrow \infty$, we define a sequence of probability measures μ_n on $\mathcal{B}(\mathcal{P})$, where

$$\mu_n(E) := \int_E h_n(q)\pi(\mathrm{d}q), \quad h_n := \frac{f_n^{m_n}}{\int_{\mathcal{P}} f_n^{m_n} \mathrm{d}\pi} = \Big(\frac{f_n}{\|f_n\|_{m_n}} \Big)^{m_n}.$$

If $\|f_n - f\|_\infty \to 0$ uniformly, with f essentially bounded), we get that the functions f_n are essentially bounded as well, thus (recall π is finite) $\int_{\mathcal{P}} f_n^{m_n} \mathrm{d}\pi < \infty$ for every n. Depending on f, we define

$$\mathcal{S}^* := \{q \in \mathcal{S} : f(q) = \|f\|_\infty\}, \quad \mathcal{S}^*_\delta := \{q \in \mathcal{S} : f(q) > \|f\|_\infty - \delta\}. \tag{4.4}$$

The following proposition is a generalization of Lember and Watkins (2022, Proposition 5.1).

Proposition 4.3. *Let π be a finite measure on $\mathcal{P}$. Let μ_n, $\mathcal{S}^*$ and $\mathcal{S}^*_\delta$ defined as above. Then for every $\delta > 0$, $\mu_n\big(\mathcal{S}^*_\delta\big) \to 1$. Moreover, if, for some $\mu \in \mathcal{P}$, $\mu\big(\bigcap_{\delta>0} \overline{\mathcal{S}^*_\delta} \setminus \mathcal{S}^*\big) = 0$ (for instance if f is upper semicontinuous) and $\mu_n \Rightarrow \mu$ then $\mu(\mathcal{S}^*) = 1$.*

Proof: Since $\|f_n\|_\infty < \infty$ and $\|f_n - f\|_\infty \to 0$ then $\|f\|_\infty < \infty$ and $\|f_n\|_\infty \to \|f\|_\infty$. Since, by assumption, π is a finite measure then for every m,

$$|\|f_n\|_m - \|f\|_m| \le \|f_n - f\|_m \le \pi(\mathcal{P})^{\frac{1}{m}} \|f_n - f\|_\infty \to 0.$$

Since $\|f\|_{m_n} \to \|f\|_\infty$, we have

$$\begin{aligned} \big|\|f_n\|_{m_n} - \|f\|_\infty\big| &\le \big|\|f_n\|_{m_n} - \|f\|_{m_n}\big| + \big|\|f\|_{m_n} - |f\|_\infty\big| \\ &\le \|f_n - f\|_{m_n} + \big|\|f\|_{m_n} - |f\|_\infty\big| \\ &\le \pi(\mathcal{P})^{\frac{1}{m_n}} \|f_n - f\|_\infty + \big|\|f\|_{m_n} - \|f\|_\infty\big| \to 0. \end{aligned}$$

Now fix $\delta > 0$ and note that

$$\mathcal{S} \setminus \mathcal{S}^*_\delta = \{q : f(q) \le \|f\|_\infty - \delta\}.$$

Define $\delta' := \delta/\|f\|_\infty$. Then

$$\begin{aligned} \mathrm{esssup}_q\Big(\mathbb{1}_{\mathcal{S}\setminus\mathcal{S}^*_\delta} \frac{f_n(q)}{\|f_n\|_{m_n}}\Big) &= \mathrm{esssup}_q\Big(\mathbb{1}_{\mathcal{S}\setminus\mathcal{S}^*_\delta} \frac{f(q) + (f_n(q) - f(q))}{\|f_n\|_{m_n}}\Big) \\ &= \mathrm{esssup}_q\Big(\mathbb{1}_{\mathcal{S}\setminus\mathcal{S}^*_\delta} \frac{f(q) + (f_n(q) - f(q))}{\|f\|_\infty} \frac{\|f\|_\infty}{\|f_n\|_{m_n}}\Big) \\ &\le \mathrm{esssup}_q\Big(\mathbb{1}_{\mathcal{S}\setminus\mathcal{S}^*_\delta} \frac{f(q)}{\|f\|_\infty} \frac{\|f\|_\infty}{\|f_n\|_{m_n}}\Big) + \frac{\|f_n - f\|_\infty}{\|f_n\|_{m_n}} \le 1 - \frac{\delta'}{2}, \end{aligned}$$

provided n is big enough. Thus,

$$\mathrm{esssup}_q(\mathbb{1}_{\mathcal{S}\setminus\mathcal{S}^*_\delta} h_n(q)) \le \Big(1 - \frac{\delta'}{2}\Big)^{m_n} \to 0,$$

so that $\mu_n(\mathcal{S}^*_\delta) \to 1$.

Suppose now that $\mu(\bigcap_{\delta>0} \overline{\mathcal{S}^*_\delta} \setminus \mathcal{S}^*) = 0$ and $\mu_n \Rightarrow \mu$. Given any $\rho \in \mathcal{P}$ and $A \subset \mathcal{P}$, define $d(\rho, A) := \inf_{\nu\in A} d(\rho, \nu)$ where d is the Prokhorov metric. It is clear that $\rho \mapsto d(\rho, A)$ is a nonnegative, continuous function such that $d(\rho, A) = 0$ if and only if $\rho \in \bar{A}$. Whence, $g_A(\rho) := \min(d(\rho, A), 1)$ is

a nonnegative, bounded continuous function.
Since $\mu_n \Rightarrow \mu$ we have

$$\int_{\mathcal{P}} g_{\mathcal{S}^*_\delta}(\rho)\mu(\mathrm{d}\rho) = \lim_{n\to+\infty} \int_{\mathcal{P}} g_{\mathcal{S}^*_\delta}(\rho)\mu_n(\mathrm{d}\rho) \leq \lim_{n\to+\infty} \mu_n(\mathcal{P} \setminus \overline{\mathcal{S}^*_\delta}) = 0,$$

whence $\mu(g_{\mathcal{S}^*_\delta} = 0) = 1$, that is $\mu(\overline{\mathcal{S}^*_\delta}) = 1$. By the continuity of the measure μ and $\mu\big(\bigcap_{\delta>0} \overline{\mathcal{S}^*_\delta} \setminus \mathcal{S}^*\big) = 0$, we have $\mu(\mathcal{S}^*) = 1$. □

Roughly speaking, according to Proposition 4.3, in order to have $\mu(\mathcal{S}^*) = 1$, we need the weak convergence $\mu_n \Rightarrow \mu$. We shall now show that under some mild conditions, when $\mathcal{S}^*$ consists of one measure q^* then the weak convergence of μ_n follows from $\mu(\mathcal{S}^*) = 1$ and the limit is δ_{q^*}. In the following corollary, let $B(q^*, \varepsilon)$ be an open ball in $\mathcal{P}$ centered at q^* and having radius ε.

Corollary 4.4. *Let the assumptions of Proposition 4.3 hold. Suppose $\mathcal{S}^* = \{q^*\}$. If, for any $\varepsilon > 0$ there exists $\delta > 0$ such that*

$$B(q^*, \varepsilon) \supseteq \mathcal{S}^*_\delta, \tag{4.5}$$

then $\mu_n \Rightarrow \delta_{q^*}$.

Proof: By hypothesis, for every $\varepsilon > 0$ there exists $\delta > 0$ so that

$$\mu_n(B(q^*, \varepsilon)) \geq \mu_n(\mathcal{S}^*_\delta).$$

By Proposition 4.3, $\mu_n(\mathcal{S}^*_\delta) \to 1$; whence, for every $\varepsilon > 0$, $\mu_n(B(q^*, \varepsilon)) \to 1$. This implies easily that $\mu_n \Rightarrow \delta_{q^*}$. □

Finally, here is an elementary but useful lemma that we write for the sake of completeness.

Lemma 4.5. *Let $\{\pi_n\}_{n\in\mathbb{N}}$ a sequence of measures on $\mathcal{P}$ weakly convergent to a measure π. Let $\{f_n\}_{n\in\mathbb{N}}$ a sequence of bounded continuous functions on $\mathcal{P}$ such that $\lim_{n\to+\infty} \|f_n - f\|_\infty = 0$ for some f. Define $\nu_n(A) := \int_A f_n(p)\pi_n(\mathrm{d}p)$ and $\nu(A) := \int_A f(p)\pi_n(\mathrm{d}p)$ for every Borel set $A \subseteq \mathcal{P}$; then $\nu_n \Rightarrow \nu$.*

Proof: Fix a bounded continuous function g on $\mathcal{P}$.

$$\begin{aligned}
\Big|\int_{\mathcal{P}} g(p)\nu_n(\mathrm{d}p) - \int_{\mathcal{P}} g(p)\nu(\mathrm{d}p)\Big| &= \Big|\int_{\mathcal{P}} g(p)f_n(p)\pi_n(\mathrm{d}p) - \int_{\mathcal{P}} g(p)f(p)\pi(\mathrm{d}p)\Big| \\
&\leq \int_{\mathcal{P}} g(p)|f_n(p) - f(p)|\pi_n(\mathrm{d}p) + \Big|\int_{\mathcal{P}} g(p)f(p)\pi_n(\mathrm{d}p) - \int_{\mathcal{P}} g(p)f(p)\pi(\mathrm{d}p)\Big| \qquad (4.6) \\
&\leq \|g\|_\infty \cdot \|f_n - f\|_\infty + \Big|\int_{\mathcal{P}} g(p)f(p)\pi_n(\mathrm{d}p) - \int_{\mathcal{P}} g(p)f(p)\pi(\mathrm{d}p)\Big| \to 0
\end{aligned}$$

as $n \to +\infty$, since $\|f_n - f\|_\infty \to 0$ and $\pi_n \Rightarrow \pi$ by hypothesis (note that $g \cdot f$ is bounded and continuous). □

The result. In this section, we consider a continuous ϕ having a unique minimum x_o. Thus, in what follows,

$$x_o := \arg\inf_x \phi(x), \quad \phi_o := \phi(x_o) = \inf_x \phi(x).$$

The following assumption is natural and assumes that if $\phi(x)$ is slightly bigger than the minimum ϕ_o, then x must be close to x_o. It also guarantees that x_o is the unique minimum and the convergence $\phi(x_n) \to \phi_o$ implies $x_n \to x_o$.

Assumption 4.6. For every $\varepsilon > 0$ there exist $\delta > 0$ so that $\{x : \phi(x) - \phi_o \leq \delta\} \subset B(x_o, \varepsilon)$, where $B(x_o, \varepsilon)$ is an open ball centered in x_o and having radius ε.

Observe that if ϕ is continuous and $\mathcal{X}$ compact, then Assumption 4.6 holds, since the minimum is unique.

Theorem 4.7. *Let the fitness function be as in (1.7) where ϕ is non-negative, continuous, bounded and satisfies Assumption 4.6; and let $\lambda \in [0,1)$. Suppose $\pi_n = \pi$ and that the support of π contains the measure δ_{x_o}. Then $\bar{\pi}_n \Rightarrow \delta_{q^*}$, where $q^* = \delta_{x_o}$. Moreover convergence (3.1) holds with the limit process being degenerate and having one almost sure path $(x_o, x_o, \ldots)$. If, in addition, $\mathcal{X}$ is compact, then $Q_n \Rightarrow \delta_{q^*}$.*

Proof: Let us start with the case $\lambda \in (0,1)$ and define

$$f_n(q) = \langle \exp[-\frac{\phi}{n^\lambda}], q\rangle^{n^\lambda}, \quad f(q) = \exp[-\langle \phi, q\rangle].$$

Since ϕ is continuous and bounded, we obtain that $q \mapsto \langle \phi, q\rangle$ is continuous and so is f. Also f_n is continuous for every n. Assuming that ϕ is bounded above, we obtain by Proposition 4.1 that for $\lambda > 0$, $\|f_n - f\|_\infty \to 0$. Since $\lambda \in (0,1)$, we take $m_n = n^{1-\lambda}$. Then

$$h_n(q) := \frac{\langle \exp[-\frac{\phi}{n^\lambda}], q\rangle^n}{Z_n} = \frac{f_n^{m_n}(q)}{Z_n}$$

so that $\mu_n = \bar{\pi}_n$. Recall that $\mathcal{S}$ is the support of $\bar{\pi}_n$. By definition

$$\mathcal{S}^* = \arg\max_{q\in\mathcal{S}} f(q) = \arg\min_{q\in\mathcal{S}} \langle \phi, q\rangle.$$

Since x_o is the unique minimum of ϕ and $q^* \in \mathcal{S}$, then $\mathcal{S}^* = \{q^*\}$ and

$$\mathcal{S}^*_\delta = \{q \in \mathcal{S}^* : \langle \phi, q\rangle - \phi_o < g(\delta)\}, \quad \text{where } g(\delta) := -\ln(1 - \delta\exp(\phi_o)).$$

Proposition 4.3 applies and so for every $\delta > 0$, it holds $\bar{\pi}_n(\mathcal{S}^*_\delta) \to 1$. In order to apply Corollary 4.4, we have to check (4.5). By the definition of Prokhorov metric, the ball $B(q^*, \varepsilon)$ consist of all measures q such that outside the ε-ball $B(x_o, \varepsilon)$ the measure q has mass at most ε, so that

$$B(q^*, \varepsilon) = \{q \in \mathcal{P} : q\big(B(x_o, \varepsilon)\big) \geq 1 - \varepsilon\}.$$

Therefore, we have to show the following: for every $\varepsilon > 0$, there exists $\delta > 0$ such that if a measure q is such that $\langle \phi, q\rangle - \phi_o \leq g(\delta)$, then it must hold that $q\big(B(x_o,\varepsilon)\big) \geq 1 - \varepsilon$. Let $\varepsilon > 0$ be fixed. By Assumption 4.6, there exists $\delta_o = \delta_o(\varepsilon) > 0$ so that $\{x : \phi(x) - \phi_o \leq \delta_o\} \subset B(x_o, \varepsilon) =: B$. Take δ such that $g(\delta)/\delta_o \leq \varepsilon$. Suppose now that the measure q satisfies $\langle \phi, q\rangle - \phi_o \leq g(\delta)$. Then

$$g(\delta) \geq \langle \phi, q\rangle - \phi_o = \int_{\mathcal{X}} (\phi - \phi_o)\mathrm{d}q \geq \int_{B^c} (\phi - \phi_o)\mathrm{d}q \geq \delta_o(1 - q(B)),$$

whence $q(B) \geq 1 - g(\delta)/\delta_o = 1 - \varepsilon$. By Corollary 4.4, $\bar{\pi}_n \Rightarrow \delta_{q^*}$.

Consider the case $\lambda = 0$. Take $f_n(q) = f(q) = \langle e^{-\phi}, q\rangle$. Since ϕ is continuous, then f is continuous. The uniform convergence is trivial and Proposition 4.3 applies. As previously, $\mathcal{S}^* = \{q^*\}$. We now have

$$\mathcal{S}^*_\delta = \{q \in \mathcal{S} : \langle e^{-\phi}, q\rangle > e^{-\phi_o} - \delta\} = \{q \in \mathcal{S} : \langle w, q\rangle > w_o - \delta\}, \quad w_o := e^{-\phi_o}.$$

Observe that Assumption 4.6 implies that the same holds for w: for every $\varepsilon > 0$ small enough there exist $\delta > 0$ so that $\{x : w(x) - w_o \leq \delta\} \subset B(x_o, \varepsilon)$. Hence, as before, (4.5) holds, and so $\bar{\pi}_n \Rightarrow \delta_{q^*}$. Thanks to Corollary 3.2, the convergence $\bar{\pi}_n \Rightarrow \delta_{q^*}$ implies that all the assumptions of Theorem 3.1 are fulfilled and so the convergence (3.1) holds and the limit process is such that with the limit only the fittest genotype survives.

For the convergence $Q_n \Rightarrow \delta_{q^*}$, we apply Theorem 3.4 and Corollary 3.2: when $\lambda \in (0,1)$, then $w \equiv 1$, and so $Q_n \Rightarrow \delta_{q^*}$. When $\lambda = 0$, then $w_n = w = e^{-\phi}$. Since ϕ is bounded, then w is bounded away from 0. Recall that $r : \mathcal{P} \to \mathcal{P}$ is as follows $r_q(A) = \langle w, q\rangle^{-1} \int_A w\mathrm{d}q$. Thus $r_{q^*} = \delta_{x_o} = q^*$. Since $\bar{\pi} = \delta_{q^*}$, it holds that $Q = \bar{\pi} r^{-1} = \delta_{r_{q^*}} = \delta_{q^*}$. □

5. Dirichlet process priors

In this section we consider the Dirichlet process priors as follows: π_n is the distribution of the Dirichlet process $D(\alpha_n)$, where α_n is a finite measure on $\mathcal{B}(\mathcal{X})$ (the base measure). Recall that a random measure P on $(\mathcal{X}, \mathcal{B}(X))$ possesses a Dirichlet process distribution $D(\alpha_n)$, when for every finite measurable partition $A_1, \dots, A_k$ of $\mathcal{X}$,

$$\big(P(A_1), \dots, P(A_k)\big) \sim \mathrm{Dir}\big(k; \alpha_n(A_1), \dots, \alpha_n(A_k)\big),$$

where $\mathrm{Dir}\big(k; \alpha_n(A_1), \dots, \alpha_n(A_k)\big)$ stands for the k-dimensional Dirichlet distribution with parameters $\big(\alpha_n(A_1), \dots, \alpha_n(A_k)\big)$. As it is common, we write $m_n := \alpha_n(\mathcal{X})$ for the total mass of the base measure, and define a probability measure $\bar{\alpha}_n := \alpha_n/m_n$. In what follows, we assume that $\bar{\alpha}_n$ is fixed and independent of n, thus $\bar{\alpha}_n = \bar{\alpha}$, but m_n depends on n, and is typically increasing in n. We also use the parallel notation $DP(m_n, \bar{\alpha})$. Since for any $A \in \mathcal{B}(\mathcal{X})$, $E[P(A)] = \bar{\alpha}(A)$ and $\mathrm{Var}[P(A)] = \frac{\bar{\alpha}(A)(1-\bar{\alpha}(A))}{m_n+1}$, we see that the bigger m_n, the more is the process concentrated on its mean $\bar{\alpha}$. Therefore increasing m_n means increasing the influence of the prior. Considering the Dirichlet process as a random element on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, i.e. $P : \Omega \to \mathcal{P}$, we define the measure π_n as its distribution $\pi_n(E) := \mathbf{P}(P \in E)$, $E \in \mathcal{B}(\mathcal{P})$. We shall refer to that prior as the Dirichlet process prior $DP(\alpha_n)$ or $DP(m_n, \bar{\alpha})$ (see equation (1.2)).

5.1. *Fixed mutation probability and fixed fitness ($\lambda = 0$).* Let us now consider the case where $m_n = c \cdot n$ where $c > 0$. Recall that c determines the mutation probability $\frac{c}{c+1}$. We assume that the support of $\bar{\alpha}$ is $\mathcal{X}$, and then also the support of π_n is $\mathcal{P}$. Recall that in this case the fitness function (1.7) is $w(x) = e^{-\phi(x)}$, where ϕ is nonnegative and continuous, (hence bounded when $\mathcal{X}$ is compact). We also assume the existence of x_o such that $\phi(x_o) = \inf_{x\in\mathcal{X}} \phi(x) =: \phi_o$. It means that $w(x_o) = \sup_{x\in\mathcal{X}} w(x)$. For the time being, x_o need not to be unique; sometimes the uniqueness is needed, and then we specify it later.

By Theorem 3.1 and Corollary 3.2, to prove convergence of P_n we just need to prove convergence of $\bar{\pi}_n$ (which was defined in (2.6)). The main idea is to prove that the sequence $\bar{\pi}_n$ satisfies a LDP with a certain rate function $I(q)$ (whence the need of compactness which is a usual assumption in large deviations theory). The rate function can be written as $I(q) = \sup_{q'\in\mathcal{P}} F(q') - F(q)$. The expression of F is different in the case $\lambda = 0$ and $\lambda > 0$, but in both cases we prove that if w has a unique maximizer, so has F. Then $I(q)$ is positive, but for one measure q^* for which $I(q^*) = 0$ and from the LDP we get convergence of $\bar{\pi}_n$ to δ_{q^*}. The details are rather lengthy and have been placed in Section 6.

In what follows, for any measurable positive function f, we shall write $\langle w, f\rangle = \int_{\mathcal{X}} wf \mathrm{d}\bar{\alpha}$ instead of using the cumbersome notation $\langle w, \mu_f\rangle$ where $\frac{\mathrm{d}\mu_f}{\mathrm{d}\bar{\alpha}} = f$.

Lemma 5.1. *If the following inequality holds:*

$$\int_{\mathcal{X}} \frac{w(x_o)}{w(x_o) - w(x)} \bar{\alpha}(\mathrm{d}x) \geq \frac{1+c}{c}, \tag{5.1}$$

then there exists only one $\theta \geq \frac{w(x_o)}{c+1}$ such that

$$f(x) := \frac{c}{\big(1 + c - \frac{w(x)}{\theta}\big)} \tag{5.2}$$

is a probability density with respect to $\bar{\alpha}$. This unique θ satisfies the (implicit) equation $\theta = \langle w, f\rangle$.

Proof: Clearly $f(x) \geq 0$ if and only if $\theta \geq \frac{w(x_o)}{c+1} =: \theta_o$. Denoting f by f_θ, we see that $\theta \mapsto \int_{\mathcal{X}} f_\theta \mathrm{d}\bar{\alpha}$ is continuous and strictly decreasing. Hence, there exists at most one θ such that f_θ is a probability density. Such θ exists, if $\int_{\mathcal{X}} f_{\theta_o}\mathrm{d}\bar\alpha \geq 1$, which is equivalent to (5.1). To see that $\theta = \langle w, f\rangle$, observe that, when f_θ is a probability density,

$$c = \int_{\mathcal{X}} \big(1 + c - \frac{w}{\theta}\big) f_\theta \mathrm{d}\bar\alpha = 1 + c - \frac{\int w f_\theta \mathrm{d}\bar\alpha}{\theta},$$

which is possible only if $\theta = \langle w, f_\theta\rangle$. □

It turns out that the inequality (5.1) is crucial. It does not hold, when w has a very sharp peak around its maximum value or $\bar\alpha$ puts very little mass around the maximum of w. Hence (5.1) somehow characterizes w as well as $\bar\alpha$. Observe that when (5.1) fails, then $\bar\alpha(w^{-1}(w(x_o)) = 0$; in particular, $\bar\alpha$ cannot have an atom at x whenever $w(x) = w(x_o)$.

In order to state and prove our main results we define two subsets of $\mathcal{P}$, namely

$$\mathcal{P}_1 := \{q \in \mathcal{P} : \bar\alpha \ll q\}, \quad \mathcal{P}_o := \{q \in \mathcal{P}_1 : q \ll \bar\alpha\}.$$

Note that, $q \in \mathcal{P}_o$ if and only if there exists a measurable function h such that $q(h > 0) = 1$ and $\mathrm{d}\bar\alpha/\mathrm{d}q = h$. In this case $q(E) = 0$ if and only if $\bar\alpha(E) = 0$; moreover $\mathrm{d}q/\mathrm{d}\bar\alpha = 1/h$.

We will see that the asymptotic distribution has a different shape according to whether or not equation (5.1) holds. On the one hand, when (5.1) holds we have a probability density given by equation (5.2) and we denote by q^* the corresponding measure; clearly $q^* \in \mathcal{P}_0$. On the other hand, when (5.1) fails, suppose that x_o is one of the absolute maxima of w (right now, we do not assume x_o to be unique), then $\bar\alpha(x_o) = 0$; since (5.1) fails, it holds

$$\beta := \int_{\mathcal{X}} f \mathrm{d}\bar\alpha < 1, \quad \text{where} \quad f := \frac{c w(x_o)}{(1 + c)(w(x_o) - w(x))} \tag{5.3}$$

and, in this case, we define the measure

$$q^* := \beta q^a + (1 - \beta)\delta_{x_o}, \tag{5.4}$$

where q^a has density $\beta^{-1} f$ (with respect to $\bar\alpha$); we shall argue in Section 6 (after (6.3)) that in this case $q^* \in \mathcal{P}_1 \setminus \mathcal{P}_o$.

Our goal is to prove that if x_o is the unique maximum for w then $\bar\pi_n \Rightarrow \delta_{q^*}$. Recall that in our case

$$\bar\pi_n(E) = \frac{1}{Z_n}\int_E \langle w, q\rangle^n \pi_n(\mathrm{d}q) = \frac{1}{Z_n}\int_E \exp[n \ln\big(\langle w, q\rangle\big)]\pi_n(\mathrm{d}q) = \frac{1}{Z_n}\int_E \exp[m_n G(q)]\pi_n(\mathrm{d}q), \tag{5.5}$$

where $G(q) := \frac{1}{c}\ln(\langle w, q\rangle)$ and $m_n := c \cdot n$. When w is bounded and continuous, then G is a continuous function on $\mathcal{P}$. We use the following theorem:

Theorem 5.2. (Feng, 2010, Corollary 9.3) *Let $\mathcal{X}$ be compact, $G : \mathcal{P} \to \mathbb{R}$ be a continuous function, and let π_n be the $DP(m_n, \bar\alpha)$-prior. Define the sequence of measures*

$$\bar\pi_n(\mathrm{d}q) = \frac{1}{Z_n}\exp[m_n G(q)]\pi_n(\mathrm{d}q).$$

The sequence satisfies a Large Deviation Principle (in short LDP) on the space $\mathcal{P}$ as n tends to infinity, with speed m_n^{-1} and rate function

$$I(q) = \sup_{q' \in \mathcal{P}}[G(q') - D(\bar\alpha\|q')] - (G(q) - D(\bar\alpha\|q))$$

where $D(\bar\alpha|q)$ is defined in equation (3.5).

Note that the following theorem holds not only with the weights w as in (1.7) but also for general continuous (hence bounded), non negative weights.

Theorem 5.3. *Let $\mathcal{X}$ be compact and let x_o be the unique maximum for w If (5.1) holds, define q^* as the measure $f\mathrm{d}\bar\alpha$, where f is as in (5.2); otherwise define q^* as in (5.4). Let r_{q^*} be the probability measure defined on $\mathcal{B}(\mathcal{X})$, such that $r_{q^*}(A) \propto \int_A w\mathrm{d}q^*$, for all $A \in \mathcal{B}(\mathcal{X})$. Then*

(1) $\bar{\pi}_n \Rightarrow \delta_{q^*}$;
(2) *the limit process of* P_n *(in the sense of* (3.1)*) is an i.i.d. process where* $X_i \sim r_{q^*}$;
(3) $Q_n \Rightarrow \delta_{r_{q^*}}$.

See Section 6 for the proof.

Example 5.4. Take $\mathcal{X} = [0,1]$, $x_o = 0.3$, $\bar{\alpha}$ – Lebesgue measure and $\phi(x) = |x - x_o|^p$. Then (5.1)) holds, if

$$\int_0^1 \Big(\frac{e^{|x-0.3|^p}}{e^{|x-0.3|^p} - 1}\Big)\mathrm{d}x \geq \frac{c+1}{c}. \tag{5.6}$$

When c is sufficiently big and p is small enough and, i.e. $p < p^*(c)$ (for instance $p(1) \approx 0.2$), then (5.6) fails.

When $p > p^*$ (so that (5.6) holds), then there exists $\theta \in [\frac{1}{1+c}, 1]$ $w(x_o) = 1)$ so that

$$f(x) = c\Big(1 + c - \frac{e^{-|x-0.3|^p}}{\theta}\Big)^{-1}$$

would integrate to 1 over $[0,1]$. Then r_{q^*} has density

$$\frac{w(x)f(x)}{\theta} = \frac{c\exp[-|x-0.3|^p]}{\theta(1+c) - \exp[-|x-0.3|^p]} = \frac{c}{(1+c)\theta e^{|x-0.3|^p} - 1} =: f^*(x).$$

Thus, when n is big enough and $X_1, \ldots, X_n \sim P_n$, then $X_1, \ldots, X_n$ are approximately i.i.d. with density f^*.

When $p < p^*$ ((5.6) fails), then $\theta = \theta_o = 1/(c+1)$ and so the function f in (5.2) is

$$f(x) = \frac{c}{1+c}\Big(1 - \exp[-|x-0.3|^p]\Big)^{-1}, \quad \beta = \int_0^1 f(x)\mathrm{d}x < 1.$$

Thus

$$\frac{f(x)w(x)}{\theta} = \frac{c}{e^{|x-0.3|^p} - 1} =: f^a(x).$$

Hence the measure r_{q^*} is such that

$$r_{q^*}(A) = \int_A f^a(x)\mathrm{d}x + (1-\beta)(1+c)\delta_{0.3}(A),$$

i.e. it has absolutely continuous part with density f^a (integrating up to $\beta(1+c) - c$) and atom 0.3 with mass $(1-\beta)(1+c)$. Thus, when n is big enough and $X_1, \ldots, X_n \sim P_n$, then $X_1, \ldots, X_n$ are approximately i.i.d. with measure r_{q^*}.

5.2. *The case* $\lambda \in (0,1)$. The authors in Lember and Watkins (2022) consider also the case when π_n is the $DP(n^{1-\lambda}, \bar{\alpha})$-prior and the fitness function is as in (1.7), i.e. $w_n = \exp[-\frac{\phi(x)}{n^\lambda}]$, $\lambda \in (0,1)$. Thus the mutation probability is $(1 + n^\lambda)^{-1} \to 0$. When $\mathcal{X}$ is finite, then in this case the limit process is again an i.i.d process with some measure q^* that differs from the measures r_{q^*} in the case $\lambda = 0$. Somehow surprisingly, the limit measure is independent of λ. We still assume the existence of x_o such that $\phi(x_o) = \inf_{x\in\mathcal{X}} \phi(x) =: \phi_o$, so that $w(x_o) = \sup_{x\in\mathcal{X}} w(x)$.

We now take π_n as $DP(c \cdot n^{1-\lambda}, \bar{\alpha})$-prior. Then the mutation probability is $c/(n^\lambda + c)$ and we need to study the measure $\bar{\pi}_n$, namely

$$\begin{aligned}\bar{\pi}_n(E) &= \frac{1}{Z_n}\int_E \langle w_n, q\rangle^n \pi_n(\mathrm{d}q) = \frac{1}{Z_n}\int_E \exp[n\ln\langle w_n, q\rangle]\pi_n(\mathrm{d}q) \\ &= \frac{1}{Z_n}\int_E \exp[n^{1-\lambda}(n^\lambda \ln\langle w_n, q\rangle)]\pi_n(\mathrm{d}q) = \frac{1}{Z_n}\int_E \exp[m_n \cdot G_n(q)]\pi_n(\mathrm{d}q)\end{aligned} \tag{5.7}$$

where $m_n := c \cdot n^{1-\lambda}$ and $G_n(q) := \frac{1}{c} \cdot n^{\lambda} \ln\langle w_n, q\rangle = \frac{1}{c} \cdot \ln\left(\langle w_n, q\rangle^{n^{\lambda}}\right)$. We see that G_n depends on n, and so Theorem 5.2 does not immediately apply. On the other hand, when ϕ is bounded, then by Proposition 4.1

$$\langle w_n, q\rangle^{n^{\lambda}} \to \exp[-\langle \phi, q\rangle]$$

and the convergence is uniform in $q \in \mathcal{P}$. Since $q \mapsto \exp[-\langle \phi, q\rangle]$ is bounded, it follows that $G_n(q) \to G(q) := -\frac{1}{c}\langle \phi, q\rangle$ uniformly.

The following result is the analogous of Lemma 5.1; the proof is similar and is left to the reader.

Lemma 5.5. *Consider the bounded function $\phi : \mathcal{X} \to [0, \infty)$ and define $\phi_o := \inf_x \phi(x)$. If the following inequality holds:*

$$\int_{\mathcal{X}} \frac{1}{\phi(x) - \phi_o} \bar{\alpha}(\mathrm{d}x) \geq \frac{1}{c}, \tag{5.8}$$

then there exists one $\theta \in [\phi_o, \phi_o + c]$ so that

$$f(x) = \frac{c}{\phi(x) + c - \theta} \tag{5.9}$$

is a probability density with respect to $\bar{\alpha}$, and then $\theta = \langle \phi, f\rangle = \int_{\mathcal{X}} \phi(x) f(x) \bar{\alpha}(\mathrm{d}x)$.

Again, when (5.8) fails, then $\bar{\alpha}(\{x \colon \phi(x) = \phi_o\}) = 0$. As in the previous case, the limit distribution takes two completely different shapes according to whether or not equation (5.8) holds. If it holds we have a probability density given by (5.9) and we denote by q^* the corresponding measure; clearly $q^* \in \mathcal{P}_0$. If not, consider x_o such that $\phi(x_o) = \phi_o$,

$$f(x) = \frac{c}{\phi(x) - \phi_o}, \quad \beta := \int_{\mathcal{X}} f \mathrm{d}\bar{\alpha} < 1, \tag{5.10}$$

and define

$$q^* = \beta q^a + (1 - \beta)\delta_{x_o}, \tag{5.11}$$

where q^a has density $\beta^{-1} f$ with respect to $\bar{\alpha}$ and $\phi(x_o) = \phi_o$. Since $f > 0$ everywhere (due to the boundedness of ϕ), it follows $q^* \in \mathcal{P}_1 \setminus \mathcal{P}_o$. Observe that q^* is independent of λ. Hence, under (5.8), it has density (5.9) with respect to $\bar{\alpha}$. Otherwise the measure q^* has atom x_o that has mass $(1-\beta)$, where β is defined as in (5.10).

The following theorem generalizes Theorem 5.2 (see Section 6.2 for the details of the proof).

Theorem 5.6. *Let $\mathcal{X}$ be compact, $G, G_n : \mathcal{P} \to \mathbb{R}$ be continuous functions that converge uniformly: $\sup_q |G_n(q) - G(q)| \to 0$, and let π_n be $DP(m_n, \bar{\alpha})$-prior. Define the sequence of measures*

$$\bar{\pi}_n(\mathrm{d}q) = \frac{1}{Z_n} \exp[m_n G_n(q)] \pi_n(\mathrm{d}q).$$

The sequence satisfies a LDP on the space $\mathcal{P}$ as n tends to infinity, with speed m_n^{-1} and rate function

$$I(q) = \sup_{q' \in \mathcal{P}} [G(q') - D(\bar{\alpha}\|q')] - (G(q) - D(\bar{\alpha}\|q)).$$

The following is the analog of Theorem 5.3 in the case $\lambda \in (0, 1)$.

Theorem 5.7. *Let $\mathcal{X}$ be compact and ϕ continuous. Assume that x_o is the unique maximum of w. If (5.8) holds, define q^* as the measure $f \mathrm{d}\bar{\alpha}$, where f is as in (5.9); otherwise define q^* as in (5.11). Then*

1. *$\bar{\pi}_n \Rightarrow \delta_{q^*}$;*
2. *the limit process of P_n (in the sense of (3.1)) is an i.i.d. process where $X_i \sim q^*$;*
3. *$Q_n \Rightarrow \delta_{q^*}$.*

The proof can be found in Section 6.

Example 5.8. Take $\mathcal{X} = [0,1]$, $x_o = 0.3$, $\bar{\alpha}$ – Lebesgue measure and $\phi(x) = |x - x_o|^p$. Then (5.8) holds, if

$$\int_0^1 |x - x_o|^{-p} \mathrm{d}x \geq \frac{1}{c}. \tag{5.12}$$

When it is so then limit process is iid process, $X_1, X_2, \ldots$ where X_i has density

$$f(x) = \frac{c}{|x - x_o|^p + c - \theta}, \quad 0 < \theta \leq c.$$

Otherwise $X_1, X_2, \ldots$ is iid process with

$$P(X_i \in A) = c \int_A |x - x_o|^{-p} \mathrm{d}x + (1 - \beta)\delta_{x_o}(A), \quad \text{where } \beta = c\int_0^1 |x - x_o|^{-p}\mathrm{d}x.$$

5.3. *The case* $\lambda \geq 1$. In the case $\lambda \geq 1$ we can treat separately the cases $\lambda = 1$ and $\lambda > 1$. The first is a special case of a constant sequence of priors, whence Theorem 4.2(1) applies where π is the $DP(c, \bar{\alpha})$.

Theorem 5.9. *Let the fitness function be as in* (1.7) *where* ϕ *is non-negative, measurable and bounded and suppose that* π_n *is the* $DP(cn^{1-\lambda}, \bar{\alpha})$ *with* $\lambda > 1$*. Then*

1. $\bar{\pi}_n \Rightarrow \delta_X$ *where* $X \sim \bar{\alpha}$*;*
2. *the limit process of* P_n *(in the sense of* (3.1)*) is* $(X, X, \ldots, X, \ldots)$ *where* $X \sim \bar{\alpha}$*;*
3. *if in addition* $\mathcal{X}$ *is compact then* $Q_n \Rightarrow \delta_X$ *where* $X \sim \bar{\alpha}$*.*

6. Proofs

6.1. *The case* $\lambda = 0$. In this section we collect all the technical results that we need to prove the main theorems of Section 5.1. Recall that $\mathcal{P}$ stands for the set of all probability measures on $\mathcal{B}(\mathcal{X})$ and remember the definitions

$$\mathcal{P}_1 := \{q \in \mathcal{P} : \bar{\alpha} \ll q\}, \quad \mathcal{P}_o := \{q \in \mathcal{P}_1 : q \ll \bar{\alpha}\}.$$

Thus $\mathcal{P}_o \subset \mathcal{P}_1$. We define the objective function F on $\mathcal{P}_1$ as follows:

$$F(q) := \ln\langle w, q\rangle - cD(\bar{\alpha}\|q),$$

where the relative entropy $D(\bar{\alpha}\|q)$ is defined by equation (3.5). We now observe that on the set $\mathcal{P}_o$,

$$F(q) = \ln\langle w, g\rangle + c\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha}, \quad \text{where} \quad g = \frac{\mathrm{d}q}{\mathrm{d}\bar{\alpha}} \quad \text{and} \quad \langle w, g\rangle = \int_{\mathcal{X}} wg\mathrm{d}\bar{\alpha}.$$

Indeed, $q \in \mathcal{P}_o$ if and only if there exists a measurable function h such that $q(h > 0) = 1$ and $\mathrm{d}\bar{\alpha}/\mathrm{d}q = h$. In this case $q(E) = 0$ if and only if $\bar{\alpha}(E) = 0$; moreover $\mathrm{d}q/\mathrm{d}\bar{\alpha} = 1/h$. It follows that when $q \in \mathcal{P}_0$ then

$$D(\bar{\alpha}\|q) = \int_{\mathcal{X}} \ln\frac{1}{g}\mathrm{d}\bar{\alpha} = -\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha}.$$

Therefore, if $q \in \mathcal{P}_0$, then

$$F(q) = \ln\big(\int_{\mathcal{X}} w\mathrm{d}q\big) - c\int_{\mathcal{X}} \ln h\mathrm{d}\bar{\alpha} = \ln\big(\int_{\mathcal{X}} wg\mathrm{d}\bar{\alpha}\big) + c\int_{\mathcal{X}} \ln g\mathrm{d}\bar{\alpha}.$$

When $q \in \mathcal{P}_o$, in order to stress the dependence of $F(q)$ on $g = \mathrm{d}q/\mathrm{d}\bar{\alpha}$, with a slight abuse of notation, we will write $F(g)$ instead of $F(q)$.

In what follows, we are interested in maximizing $F(q)$ over $\mathcal{P}_1$ and finding the argmax, when it exists. We split the maximization problem into two parts: maximizing over $\mathcal{P}_o$ and $\mathcal{P}_1 \setminus \mathcal{P}_o$. On $\mathcal{P}_o$, it holds $F(q) = F(g)$, where g is the density of q with respect to $\bar{\alpha}$. Hence

$$\sup_{q\in\mathcal{P}_o} F(q) \leq \sup_{g\in\mathcal{F}} F(g),$$

where $\mathcal{F}$ is the set of all probability densities with respect to $\bar{\alpha}$. Here there is an inequality, because the definition of $\mathcal{P}_o$ implies that $\operatorname{supp} g = \mathcal{X}$, but $\mathcal{F}$ is the set of all probability densities. We start with maximizing F over $\mathcal{F}$.

6.1.1. *Maximizing F over $\mathcal{F}$, when inequality (5.1) holds.* We now show that under (5.1) the function f as in (5.2) is the unique solution of the above-stated maximization problem. Observe that $F(f) < \infty$. Indeed, since $\frac{c}{(c+1)} \leq f(x) \leq \frac{c}{(c+1)-\frac{w(x_o)}{\theta}}$ then $\ln f(x)$ is bounded from below; moreover $\ln f(x) \leq f(x) - 1$.

Lemma 6.1. *If (5.1) holds, then $\sup_{g\in\mathcal{F}} F(g) = F(f)$ and $F(g) < F(f)$ when $g \neq f$, where f is given by (5.2).*

Proof: Let f be the density (5.2). It suffices to show that for any other $g \in \mathcal{F}$ such that $g \neq f$ $\bar{\alpha}$-a.s.

$$F(f) - F(g) = \ln\theta + c\int_{\mathcal{X}} \ln f \mathrm{d}\bar{\alpha} - \ln\theta' - c\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} = \ln\frac{\theta}{\theta'} - c\int_{\mathcal{X}} \ln\frac{g}{f}\mathrm{d}\bar{\alpha} > 0,$$

where $\theta' = \langle w, g\rangle$. When $\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} = -\infty$, then the strict inequality holds, otherwise observe that all integrals above are finite and since

$$\int_{\mathcal{X}} \frac{g}{f}\mathrm{d}\bar{\alpha} = \frac{1+c}{c}\Big(1 - \frac{\theta'}{\theta(c+1)}\Big),$$

by Jensen inequality we get

$$-c\int_{\mathcal{X}} \ln\frac{g}{f}\mathrm{d}\bar{\alpha} > -\ln\Big[\frac{1+c}{c}\Big(1-\frac{\theta'}{\theta(c+1)}\Big)\Big]^c,$$

where the strict inequality follows from assumption $f \neq g$ $\bar{\alpha}$ - a.s.. Therefore, it suffices to show that

$$\ln\frac{\theta}{\theta'} - \ln\Big[\frac{1+c}{c}\Big(1-\frac{\theta'}{\theta(c+1)}\Big)\Big]^c \geq 0.$$

The latter is equivalent to

$$\ln\Big[\frac{1+c}{c}\Big(1-\frac{\theta'}{\theta(c+1)}\Big)\Big]^c - \ln\frac{\theta}{\theta'} \leq 0 \quad \Leftrightarrow \quad \Big[\frac{1+c}{c}\Big(1-\frac{\theta'}{\theta(c+1)}\Big)\Big]^c \cdot \frac{\theta'}{\theta} \leq 1. \tag{6.1}$$

Since $\frac{w(x_o)}{c+1} \leq \theta = \langle w, f\rangle \leq w(x_o)$, and $\theta' = \langle w, g\rangle \leq w(x_o)$, it holds that

$$0 \leq \frac{\theta'}{\theta(c+1)} \leq \frac{w(x_o)}{w(x_o)} = 1.$$

Denoting $\frac{\theta'}{\theta(c+1)} =: 1-\alpha$, we obtain that the right hand side inequality in (6.1) is

$$\Big[\Big(1+\frac{1}{c}\Big)\alpha\Big]^c (c+1)(1-\alpha) \leq 1$$

and this holds by Proposition 6.2, proven below. □

Proposition 6.2.

$$\max_{x\geq 0, \alpha\in[0,1]} \Big((1+x)(1+1/x)^x\alpha^x(1-\alpha)\Big) = 1. \tag{6.2}$$

Proof: Fix $\alpha \in (0,1)$. Then $x_\alpha = \frac{\alpha}{1-\alpha}$ attains the maximum of

$$u(x) := (1+x)(1+1/x)^x\alpha^x$$

over $[0,\infty)$. To see that observe: $\lim_{x\to\infty} u(x) = 0$, $\lim_{x\to 0} u(x) = 1$ and

$$u'(x) = (\ln u(x))' u(x) = \big(\ln(1+1/x) + \ln\alpha\big)u(x),$$

so that x_α is the only stationary point of the function $u(x)$. Plugging x_α into the left hand side of (6.2), we obtain

$$(1+x_\alpha)(1+1/x_\alpha)^{x_\alpha}\alpha^{x_\alpha}(1-\alpha)=1.$$

□

6.1.2. *Maximizing F over $\mathcal{F}$, when inequality (5.1) fails.* Let x_o be such that $w(x_o)=\sup_x w(x)=:\bar{w}$. Right now, we do not assume x_o to be unique. Since (5.1) fails, we define β, f and q^* as in (5.3) and (5.4). Clearly

$$\langle w,q^*\rangle=\langle w,f\rangle+(1-\beta)\bar{w}=\bar{w}\big(\beta-\frac{c}{1+c}+1-\beta\big)=\frac{\bar{w}}{1+c}=\theta_o; \tag{6.3}$$

recall that θ_o was defined in the proof of Lemma 5.1. Observe that $q^*\in\mathcal{P}_1\setminus\mathcal{P}_o$. Indeed, since (5.1) fails, then $0=\bar{\alpha}(w^{-1}(\bar{w}))\geq\bar{\alpha}(x_o)$ while $q^*(x_o)>0$. On the other hand, since $f(x)>0$ whenever $x\notin w^{-1}(\bar{w})$, it follows that $q^*(E)=0$ implies $q^a(E)=0$ whence $\bar{\alpha}(E)=0$. In this case $\mathrm{d}\bar{\alpha}/\mathrm{d}q^*=\mathbb{1}_{\mathcal{X}\setminus\{x_o\}}/f$. Now, according to (6.1)

$$F(q^*)=\ln\theta_o+c\int_{\mathcal{X}}\ln f\mathrm{d}\bar{\alpha}. \tag{6.4}$$

Lemma 6.3. *Let (5.1) fail and $g\in\mathcal{F}$. Then $F(g)<F(q^*)$, where $F(q^*)$ is defined as in (6.4).*

Proof: Observe that $F(q^*)$ is independent of the choice of x_o. The proof is the exactly same as that of Lemma 6.1. The Jensen inequality is an equality, when f/g is constant $\bar{\alpha}$-a.s.. In our case it means $g=\beta^{-1}f$. Note that

$$\ln\big(\bar{w}(\beta-\frac{c}{1+c})\big)-(c+1)\ln\beta<\ln\big(\frac{\bar{w}}{1+c}\big)\quad\Leftrightarrow\quad\beta(1+c)-c<\beta^{c+1}$$

and for $\beta<1$ the L.H.S. holds. Thus

$$\begin{aligned}F(g)&=\ln(\langle w,f\rangle)-\ln\beta+c\int_{\mathcal{X}}\ln f\mathrm{d}\bar{\alpha}-c\ln\beta=\ln\big(\bar{w}(\beta-\frac{c}{1+c})\big)-(c+1)\ln\beta+c\int_{\mathcal{X}}\ln f\mathrm{d}\bar{\alpha}\\&<\ln\big(\frac{\bar{w}}{1+c}\big)+c\int_{\mathcal{X}}\ln f\mathrm{d}\bar{\alpha}=F(q^*).\end{aligned}$$

□

6.1.3. *Maximizing F over $\mathcal{P}_1\setminus\mathcal{P}_o$, when inequality (5.1) holds.* Let f be the density (5.2).

Lemma 6.4. *Let $q\in\mathcal{P}_1\setminus\mathcal{P}_o$. Then $F(q)<F(f)$.*

Proof: Let $h=\frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}q}$. Let $q(h>0)=:\beta_1$; thus $q(h=0)=1-\beta_1$. Since $q\notin\mathcal{P}_o$, $\beta_1<1$, because otherwise $1/h$ would be a density. Clearly $\beta_1>0$. Let $g:=1/h$. Thus $q(g=\infty)>0$, but $\bar{\alpha}(g=\infty)=\bar{\alpha}(h=0)=0$. For any Borel set B

$$\int_B g\mathrm{d}\bar{\alpha}=\int_{B\cap\{h>0\}}g\mathrm{d}\bar{\alpha}=\int_{B\cap\{h>0\}}\frac{1}{h}h\mathrm{d}q=\int_{B\cap\{h>0\}}\mathrm{d}q=q(B\cap\{h>0\}).$$

Whence

$$q(B)=\int_B g\mathrm{d}\bar{\alpha}+q(\{h=0\}\cap B). \tag{6.5}$$

We also get that

$$\int_{\mathcal{X}}g\mathrm{d}\bar{\alpha}=q(h>0)=\beta_1,\qquad\int_{\{h>0\}}w\mathrm{d}q=\int_{\{h>0\}}wg\mathrm{d}\bar{\alpha}=\int_{\mathcal{X}}wg\mathrm{d}\bar{\alpha},$$

because $\bar{\alpha}(h=0)=0$ and so

$$\theta':=\langle w,q\rangle=\int_{\{h>0\}}wg\mathrm{d}\bar{\alpha}+\int_{\{h=0\}}w\mathrm{d}q\leq\int_{\mathcal{X}}wg\mathrm{d}\bar{\alpha}+\bar{w}(1-\beta_1)=\langle w,g\rangle+\bar{w}(1-\beta_1), \tag{6.6}$$

where the equality holds if and only if $q(h=0)=q(h=0,\, w=\bar{w})$ (recall that $q(h=0)=1-\beta_1$). By equation (6.1),

$$F(q)=\ln[\langle w,q\rangle]-c\int_{\mathcal{X}}\ln h\mathrm{d}\bar{\alpha}=\ln\theta'+c\int_{\mathcal{X}}\ln g\mathrm{d}\bar{\alpha}. \tag{6.7}$$

Therefore, it suffices to prove that

$$F(f)-F(q)=\ln\theta+c\int_{\mathcal{X}}\ln f\mathrm{d}\bar{\alpha}-\ln\theta'-c\int_{\mathcal{X}}\ln g\mathrm{d}\bar{\alpha}=\ln\frac{\theta}{\theta'}-c\int_{\mathcal{X}}\ln\frac{g}{f}\mathrm{d}\bar{\alpha}>0.$$

The proof follows the steps of Lemma 6.1. In this case

$$\int_{\mathcal{X}}\frac{g}{f}\mathrm{d}\bar{\alpha}=\frac{1+c}{c}\Big(\beta_1-\frac{\langle w,g\rangle}{\theta(1+c)}\Big) \tag{6.8}$$

so that, after using the Jensen inequality, instead of the inequality (6.1), now we have

$$\Big[\frac{1+c}{c}\Big(\beta_1-\frac{\langle w,g\rangle}{\theta(1+c)}\Big)\Big]^c\frac{\theta'}{\theta}\le 1. \tag{6.9}$$

To see that (6.9) holds, define

$$\alpha:=\beta_1-\frac{\langle w,g\rangle}{\theta(1+c)}.$$

Since $\langle w,g\rangle\le\bar{w}\beta_1$ and $\theta(1+c)\ge\bar{w}$, it holds that $\alpha\in[0,\beta_1]$. On the other hand, by (6.6)

$$\frac{\theta'}{\theta}\le\frac{\langle w,g\rangle}{\theta}+\frac{\bar{w}(1-\beta_1)}{\theta}\le(\beta_1-\alpha)(1+c)+(1+c)(1-\beta_1)=(1+c)(1-\alpha).$$

Hence

$$\Big[\frac{1+c}{c}\Big(\beta_1-\frac{\langle w,g\rangle}{\theta(1+c)}\Big)\Big]^c\frac{\theta'}{\theta}\le(1+c)\big(1+\frac{1}{c}\big)^c\alpha^c(1-\alpha)\le 1,$$

where the last inequality comes from Proposition 6.2. This proves the strict inequality if the Jensen inequality is strict.

Since $\int_{\mathcal{X}}g\mathrm{d}\bar{\alpha}=\beta_1<1$, we obtain that the Jensen inequality is an equality if and only if $f=\beta_1^{-1}g$; in this case by (6.6), since $\langle w,g\rangle=\beta_1\langle w,f\rangle=\beta_1\theta$, we have $\theta'\le\theta(\beta_1+(1-\beta_1)\frac{\bar{w}}{\theta})$ and $\int_{\mathcal{X}}\ln\frac{g}{f}\mathrm{d}\bar{\alpha}=\ln(\beta_1)$ so that

$$\ln\frac{\theta}{\theta'}-c\int_{\mathcal{X}}\ln\frac{g}{f}\mathrm{d}\bar{\alpha}\ge-\ln(\beta_1+(1-\beta_1)\frac{\bar{w}}{\theta})-c\ln\beta_1\ge-\ln(\beta_1+(1-\beta_1)(1+c))-c\ln\beta_1>0,$$

because for every $c>0$, it holds

$$\beta_1^c(\beta_1+(1-\beta_1)(1+c))=\beta_1^c(1-c\beta_1+c)<1.$$

□

6.1.4. *Maximizing F over $\mathcal{P}_1\setminus\mathcal{P}_o$, when inequality (5.1) fails.*

Lemma 6.5. *Assume that x_o is unique. Let $q\in\mathcal{P}_1\setminus\mathcal{P}_o$ and $q\neq q^*$, where q^* is defined as in (5.4). Then $F(q)<F(q^*)$.*

Proof: We know that, since (5.1) fails, then $\bar{\alpha}(x_o)=0$, whence $q^*\in\mathcal{P}_1\setminus\mathcal{P}_o$ and since x_o is unique, the construction (5.4) uniquely defines q^* as well. We have to prove

$$F(q^*)-F(q)=\ln\frac{\theta_o}{\theta'}-c\int_{\mathcal{X}}\ln\frac{g}{f}\mathrm{d}\bar{\alpha}>0, \tag{6.10}$$

where

$$\theta_o=\frac{\bar{w}}{1+c},\quad h=\frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}q},\quad g=h^{-1},\quad \beta_1=\int_{\mathcal{X}}g\mathrm{d}\bar{\alpha},\quad \theta'=\langle w,q\rangle\le\langle w,g\rangle+(1-\beta_1)\bar{w}$$

and f is as in (5.3). Observe $\int_{\mathcal{X}} f \mathrm{d}\bar{\alpha} = \beta = \frac{\langle w,f\rangle}{\bar{w}} + \frac{c}{1+c} \geq \frac{c}{1+c}$. Again, we follow the steps of Lemma 6.1. Since now $\theta_o(1+c) = \bar{w}$, from (6.8) we get

$$\int_{\mathcal{X}} \frac{g}{f} \mathrm{d}\bar{\alpha} = \frac{1+c}{c}\Big(\beta_1 - \frac{\langle w,g\rangle}{\bar{w}}\Big) \tag{6.11}$$

and with $\alpha = \beta_1 - \frac{\langle w,g\rangle}{\bar{w}}$ and as in the proof of Lemma 6.4 we obtain that (6.9) holds.

When the Jensen inequality is strict, there is nothing to prove. Observe that the Jensen inequality is an equality if and only if $g = \frac{\beta_1}{\beta} f$ $\bar{\alpha}$-a.s. So, when $\beta = \beta_1$, it means that $f = g$. This implies that the inequality $\theta' \leq \langle w,g\rangle + (1-\beta_1)\bar{w} = \langle w,f\rangle + (1-\beta)\bar{w} = \theta_o$ must be strict. Indeed, otherwise by (6.6), $q(h=0) = q(h=0,\, w=\bar{w}) = 1-\beta_1$ but $w^{-1}(\bar{w}) = \{x_o\}$, thus $q(x_o) = 1-\beta_1 = 1-\beta = q^*(x_o)$ whence $q = q^*$. But when $\theta' < \theta_o$, and $g = f$, then (6.10) trivially holds.

Let us now consider the case $\beta_1 \neq \beta$ and $g = \frac{\beta_1}{\beta} f$ $\bar{\alpha}$-almost surely. In this case

$$\langle w,g\rangle = \frac{\beta_1}{\beta}\langle w,f\rangle = \frac{\beta_1}{\beta}\big(\beta - \frac{c}{1+c}\big)\bar{w}$$

so that

$$\theta' \leq \langle w,g\rangle + (1-\beta_1)\bar{w} = \bar{w}\big(1 - \frac{\beta_1}{\beta}\frac{c}{1+c}\big) = \bar{w}\big(\frac{\beta(1+c)-\beta_1 c}{\beta(1+c)}\big), \quad \ln\frac{\theta_o}{\theta'} \geq \ln\frac{\beta}{\beta(1+c)-\beta_1 c}.$$

Now

$$\ln\frac{\theta_o}{\theta'} - c\int_{\mathcal{X}} \ln\frac{g}{f}\mathrm{d}\bar{\alpha} \geq \ln\frac{\beta}{\beta(1+c)-\beta_1 c} - c\ln\frac{\beta_1}{\beta} > 0.$$

To see that the last inequality holds note that it is equivalent to

$$\beta^{c+1} > \beta_1^c(\beta(1+c)-\beta_1 c).$$

The function $\beta \mapsto \beta^{c+1} - \beta_1^c(\beta(1+c)-\beta_1 c)$ is strictly positive at $\frac{c}{c+1}$ and at 1 (both statements hold for any $\beta_1 \in [0,1]$ and $c > 0$). The function has unique minimum at β_1, where it equals 0 and hence it is strictly positive elsewhere. □

6.1.5. *Proof of the main theorem when $\lambda = 0$.*

Proof of Theorem 5.3: (1) For any $q \in \mathcal{P}_1$, $\frac{1}{c}\ln(\langle w,q\rangle) - D(\bar{\alpha}\|q) = \frac{1}{c}F(q)$. Since for $q \notin \mathcal{P}_1$ we have $D(\bar{\alpha}\|q) = \infty$, then

$$\sup_{q\in\mathcal{P}}\Big[\frac{1}{c}\ln(\langle w,q\rangle) - D(\bar{\alpha}\|q)\Big] = \sup_{q\in\mathcal{P}_1}\frac{1}{c}F(q) = \frac{1}{c}F(q^*).$$

Indeed, if (5.1) holds, the last equality follows from Lemma 6.1 and Lemma 6.4. On the other hand, if (5.1) fails, then we apply Lemma 6.3 and Lemma 6.5 instead. By assumption x_o is unique so $q^* \in \mathcal{P}_1$ is the unique maximizer of $F(q)$. Therefore

$$I(q) = \begin{cases} \frac{1}{c}\big(F(q^*) - F(q)\big), & \text{if } q \in \mathcal{P}_1; \\ \infty, & \text{else.} \end{cases}$$

The LDP implies: for any closed set C:

$$\limsup_n \frac{1}{cn}\ln\bar{\pi}_n(C) \leq -\inf_{q\in C} I(q). \tag{6.12}$$

By Lemma 6.2.13 in Dembo and Zeitouni (2010),

$$D(\bar{\alpha}\|q) = \sup_{g\in C_b}[\langle g,\bar{\alpha}\rangle - \ln\langle e^g,q\rangle].$$

For every g, the function $q \mapsto \langle g,\bar{\alpha}\rangle - \ln\langle e^g,q\rangle$ is continuous and so their supremum $D(\bar{\alpha}\|\cdot)$ is lower semicontinuous. Therefore F is a upper semicontinuous function. Now take $C = B^c$, where B is an open ball (with respect to the Prokhorov metric) containing q^*. Thus, C is compact and so $\sup_{q\in C} F(q) = F(q_o)$ for some $q_o \in C$. Moreover we know that $F(q) < F(q^*)$ for every $q \neq q^*$, in particular $F(q_o) < F(q^*)$ since $q^* \notin C$. So we have shown that

$$\sup_{q\in C} F(q) < F(q^*). \tag{6.13}$$

From (6.13) we have $\inf_{q\in C} I(q) > 0$ and so from (6.12), it follows that $\bar\pi_n(C) \to 0$ exponentially fast. This means $\bar\pi_n \Rightarrow \delta_{q^*}$.

(2) When $\lambda = 0$, $w_n = w$ and so $r_{q,n} = r_q$. By Theorem 3.1 and Corollary 3.2, the limit process $X_1, X_2, \ldots$ exists and it is the i.i.d process, where $X_i \sim r_{q^*}$. When (5.1) holds, then

$$r_{q^*}(A) = \frac{1}{\langle w, f\rangle}\int_A w(x)f(x)\bar\alpha(\mathrm{d}x), \quad A \in \mathcal{B}(\mathcal{X}),$$

where f is as in (5.2). When (5.1) fails, then (recall (6.3))

$$r_{q^*}(A) = \frac{1}{\theta_o}\Big(\int_A w(x)f(x)\bar\alpha(\mathrm{d}x) + (1-\beta)w(x_o)\delta_{x_o}(A)\Big), \quad A \in \mathcal{B}(\mathcal{X}),$$

where f is as in (5.3) and $\beta = \int_{\mathcal{X}} f\mathrm{d}\bar\alpha$. Since $\theta_o = \frac{w(x_o)}{1+c}$, we see that when (5.1) fails then $\bar\alpha(\{x\colon w(x) = w(x_o)\}) = 0$, hence the proportion of x_o-types in the limit population equals $r_{q^*}(\{x_o\}) = (1-\beta)(1+c)$.

(3) The convergence of Q_n follows from Theorem 3.4.

□

6.2. *The case $\lambda \in (0,1)$.* This section contains the technical result that we need to prove the main theorems of Section 5.2.

6.2.1. *A generalization of Theorem 5.2.* In Section 5.2, we stated Theorem 5.6 which is a generalization of Theorem 5.2 for uniform convergence. Our first task now is to prove this theorem. The proof relies on Theorems 6.6, 6.7 and 6.8 below.

Theorem 6.6. (Feng, 2010, Theorem 9.2) *Assume $\mathcal{X}$ is compact. Then π_n satisfies the LDP with speed m_n^{-1} and rate function $H(q) = D(\bar\alpha\|q)$.*

Here the rate function $H(q)$ equals $D(\bar\alpha\|q)$. As argued in the proof of Theorem 5.3, $D(\bar\alpha\|\cdot)$ is lower semicontinuous and so the level set $\{q : D(\bar\alpha\|q) \leq \alpha\}$ is closed for every $\alpha > 0$. Recall that a rate function is good if all level sets are compact; hence if $\mathcal{X}$ is compact, then the rate function $H(q) := D(\bar\alpha\|q)$ is good.

Theorem 6.7. (Varadhan lemma, Feng, 2010, Theorem B.1) *Assume $\mathcal{X}$ is compact and π_n satisfies the LDP with speed m_n and good rate function H. Let G_n and G be a family of continuous functions on $\mathcal{P}$ satisfying $\sup_q |G_n(q) - G(q)| \to 0$ Then*

$$\lim_n \frac{1}{m_n}\ln\int_{\mathcal{P}}\exp[m_n\cdot G_n(q)]\pi_n(\mathrm{d}q) = \sup_q\big(G(q) - H(q)\big). \tag{6.14}$$

Theorem 6.8. (Feng, 2010, Theorem B.6)) *Assume $\mathcal{X}$ is compact and $\bar\pi_n$ is such that*

$$\lim_{\delta\to 0}\limsup_n \frac{1}{m_n}\ln\bar\pi_n\big(B(q,\delta)\big) = \lim_{\delta\to 0}\liminf_n\frac{1}{m_n}\ln\bar\pi_n\big(B^o(q,\delta)\big) = -I(q),$$

where $B(q,\delta) := \{p : d(p,q) \leq \delta\}$, $B^o(q,\delta) := \{p : d(p,q) < \delta\}$ and I is a good rate function. Then $\bar\pi_n$ satisfies the LDP with rate function I and speed m_n^{-1}.

Proof of Theorem 5.6: Since $\sup_q |G_n(q) - G(q)| \to 0$, for every $\varepsilon_1 > 0$ there exists n_1 so that $|G_n(q) - G(q)| \le \varepsilon_1$, whenever $n > n_1$. Fix q and ε_2 and take $\delta > 0$ so small $|G(q) - G(p)| \le \varepsilon_2$ for every $p \in B(q,\delta) =: B$. Estimate

$$\int_B e^{m_n G_n(p)} \pi_n(dp) \le \int_B e^{m_n \left(G(p)+\varepsilon_1\right)} \pi_n(dp) \le \int_B e^{m_n \left(G(q)+\varepsilon_1+\varepsilon_2\right)} \pi_n(dp) = e^{m_n \left(G(q)+\varepsilon_1+\varepsilon_2\right)} \pi_n(B).$$

Then by Theorem 6.6

$$\begin{aligned}\limsup_n \frac{1}{m_n} \ln \Big(\int_B e^{m_n G_n(p)} \pi_n(dp) \Big) &\le \big(G(q) + \varepsilon_1 + \varepsilon_2\big) + \limsup_n \big(\ln \pi_n(B)\big) \\ &\le G(q) + \varepsilon_1 + \varepsilon_2 - \inf_{p\in B} H(p).\end{aligned}$$

Similarly, with $B^o = B^o(q,\delta)$,

$$\begin{aligned}\liminf_n \frac{1}{m_n} \ln \Big(\int_B e^{m_n G_n(p)} \pi_n(dp) \Big) &\ge \big(G(q) - \varepsilon_1 - \varepsilon_2\big) + \liminf_n \big(\ln \pi_n(B^o)\big) \\ &\ge G(q) - \varepsilon_1 - \varepsilon_2 - \inf_{p\in B^o} H(p).\end{aligned}$$

Since

$$Z_n = \int_{\mathcal{P}} e^{m_n G_n(p)} \pi_n(dp),$$

by Theorem 6.7,

$$\frac{1}{m_n} \ln Z_n \to \sup_p \big(G(p) - H(p)\big).$$

Therefore

$$\begin{aligned}&\limsup_n \frac{1}{m_n} \ln \bar{\pi}_n(B) \le G(q) + \varepsilon_1 + \varepsilon_2 - \inf_{p\in B} H(p) - \sup_p \big(G(p) - H(p)\big) \\ &\liminf_n \frac{1}{m_n} \ln \bar{\pi}_n(B^o) \ge G(q) - \varepsilon_1 - \varepsilon_2 - \inf_{p\in B^o} H(p) - \sup_p \big(G(p) - H(p)\big)\end{aligned}$$

Let $\delta \to 0$. Then $\varepsilon_2(\delta)$ goes to 0 and $\lim_n - \inf_{p\in B} H(p) = -H(p)$, because $-H$ is upper semi-continuous and so

$$\lim_{\delta\to 0} - \inf_B H(p) = -H(q)$$

because $-\inf_{p\in B(q,\delta_n)} H(p) \ge -H(q)$ and so, taking $\delta_n \to 0$, $\liminf_n - \inf_{p\in B(q,\delta_n)} H(p) \ge -H(q)$, but when $p_n \Rightarrow q$ is such that $\limsup_n - \inf_{p\in B(q,\delta_n)} H(p) = \limsup_n -H(p_n)$, then by USC, it holds $\limsup_n - \inf_{p\in B_n} H(p) \le -H(q)$. The same holds when B is replaced by B^o.
Therefore

$$\begin{aligned}&\lim_{\delta\to 0} \limsup_n \frac{1}{m_n} \ln \bar{\pi}_n(B) \le G(q) + \varepsilon_1 - H(q) - \sup_p \big(G(p) - H(p)\big) \\ &\lim_{\delta\to 0} \liminf_n \frac{1}{m_n} \ln \bar{\pi}_n(B^o) \ge G(q) - \varepsilon_1 - H(q) - \sup_p \big(G(p) - H(p)\big)\end{aligned}$$

Since ε_1 was arbitrary, we see that the assumptions of Theorem 6.8 hold, therefore $\bar{\pi}_n$ satisfies the LDP with speed m_n and rate function I. □

6.2.2. *The objective function F.* We now define

$$F(q) = -\langle \phi, q\rangle - cH(q), \quad q \in \mathcal{P}_1.$$

When $q \in \mathcal{P}_0$, then $\exists g = \frac{dq}{d\bar{\alpha}}$ and then

$$F(q) = -\int_{\mathcal{X}} \phi g d\bar{\alpha} + c\int_{\mathcal{X}} \ln g d\bar{\alpha} = -\langle \phi, q\rangle + c\int_{\mathcal{X}} \ln g d\bar{\alpha} = F(g).$$

6.2.3. *Maximizing F, when inequality (5.8) holds.*

Lemma 6.9. *Assume that (5.8) holds. Then F has a unique maximizer over $\mathcal{P}_1$, which is the measure with density (5.9) with respect to $\bar{\alpha}$.*

Proof: Let f be the density (5.9). Again, we split the maximization: over $\mathcal{F}$ and over $\mathcal{P}_1 \setminus \mathcal{P}_0$. At first, we show that for every $g \in \mathcal{F}$ such that $g \neq f$ $\bar{\alpha}$-a.s., it holds

$$F(f) - F(g) = -\theta + c\int_{\mathcal{X}} \ln f \mathrm{d}\bar{\alpha} + \theta' - c\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} = \theta' - \theta - c\int_{\mathcal{X}} \ln \frac{g}{f} \mathrm{d}\bar{\alpha} > 0,$$

where $\theta' = \langle \phi, g\rangle$. When $\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} = -\infty$, then inequality strictly holds, otherwise observe that all integrals above are finite and since

$$\int_{\mathcal{X}} \frac{g}{f} \mathrm{d}\bar{\alpha} = \frac{\theta' - \theta + c}{c},$$

by Jensen inequality we get

$$-c\int_{\mathcal{X}} \ln \frac{g}{f} \mathrm{d}\bar{\alpha} > -\ln\Big[\frac{\theta' - \theta + c}{c}\Big]^c,$$

where the strict inequality follows from assumption $f \neq g$ $\bar{\alpha}$ - a.s.. Therefore, it suffices to show that the L.H.S. is nonnegative; to this aim, note that

$$\theta' - \theta - \ln\Big[\frac{\theta' - \theta + c}{c}\Big]^c = c\Big(\frac{\theta' - \theta}{c} - \ln\Big[\frac{\theta' - \theta}{c} + 1\Big]\Big) \geq 0.$$

We now take $q \in \mathcal{P}_1 \setminus \mathcal{P}_o$. Let $h = \frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}q}$, $g = 1/h$, $\beta_1 = \int_{\mathcal{X}} g \mathrm{d}\bar{\alpha} \equiv q(h > 0)$, $\theta' = \langle \phi, q\rangle \geq \langle \phi, g\rangle + \phi_o(1 - \beta_1)$. As previously, we obtain via Jensen's inequality

$$F(f) - F(q) = -\theta + c\int_{\mathcal{X}} \ln f \mathrm{d}\bar{\alpha} + \theta' - c\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} \geq \theta' - \theta - c\ln\big(\frac{\langle \phi, g\rangle + (c - \theta)\beta_1}{c}\big) \tag{6.15}$$

and it remains to show that

$$\ln\big(\frac{\langle \phi, g\rangle + (c - \theta)\beta_1}{c}\big) \leq \frac{\theta' - \theta}{c}.$$

Since $\langle \phi, g\rangle \leq \theta' - \phi_o(1 - \beta_1)$ and $c - \theta \geq -\phi_o$ we obtain that

$$\begin{aligned}\ln\big(\frac{\langle \phi, g\rangle + (c - \theta)\beta_1}{c}\big) &\leq \ln\big(\frac{\theta' + (c - \theta)\beta_1 - \phi_o(1 - \beta_1)}{c}\big) \leq \ln\big(\frac{\theta' + (c - \theta)\beta_1 + (c - \theta)(1 - \beta_1)}{c}\big)\\ &\leq \ln\big(1 + \frac{\theta' - \theta}{c}\big) \leq \frac{\theta' - \theta}{c}.\end{aligned}$$

Now it remains to argue that at least one of the inequalities is strict. The Jensen inequality is an equality only if $\beta_1 f = g$ and in this case $F(f) - F(q) = \theta' - \theta - c\ln\beta_1$. Since for $g = \beta_1 f$, it holds $\theta' \geq \beta_1\theta + \phi_o(1 - \beta_1)$ we obtain that $\theta' - \theta \geq (1 - \beta_1)(\phi_o - \theta) \geq -c(1 - \beta_1) > c\ln\beta_1$, and so $F(f) - F(q) = \theta' - \theta - c\ln\beta_1 > 0$. □

6.2.4. *Maximizing F, when inequality (5.8) fails.* Remember the definition of q^* given in equation (5.11). Now

$$\langle \phi, q^*\rangle = \langle \phi, f\rangle + (1 - \beta)\phi_o = c + \phi_o\beta + (1 - \beta)\phi_o = c + \phi_o =: \theta_o. \tag{6.16}$$

Since (5.1) fails then $\bar{\alpha}(x_o) = 0$, which in turn implies $q^* \in \mathcal{P}_1$, and then $F(q^*) = -\phi_o + c\int_{\mathcal{X}} \ln f d\bar{\alpha}$.

Lemma 6.10. *Let (5.8) fail, x_o be the unique minimizer of ϕ. Then for every $q \in \mathcal{P}_1$ such that $q \neq q^*$, it holds $F(q) < F(q^*)$.*

Proof: Again, we start with maximizing over $\mathcal{F}$. Let $g \in \mathcal{F}$. Then, by Jensen Inequality

$$F(q^*) - F(g) = -\theta_o + c\int_{\mathcal{X}} \ln f \mathrm{d}\bar{\alpha} + \theta' - c\int_{\mathcal{X}} \ln g \mathrm{d}\bar{\alpha} = \theta' - (\phi_o + c) - c\int_{\mathcal{X}} \ln \frac{g}{f}\mathrm{d}\bar{\alpha}$$
$$\geq \theta' - (\phi_o + c) - c\ln\Big[\frac{\theta' - \phi_o}{c}\Big] \geq 0,$$

because for every x, $x - 1 \geq \ln x$. The Jensen inequality is an equality, when $\beta g = f$ and then $F(q^*) - F(g) = \theta' - \theta_o + c\ln\beta$. Then also $\theta' = \langle\phi, g\rangle = \beta^{-1}\langle\phi, f\rangle = c/\beta + \phi_o$ and, therefore,

$$F(q^*) - F(g) = \theta' - c - \phi_o + c\ln\beta = c/\beta + \phi_o - c - \phi_o + c\ln\beta = c(\frac{1}{\beta} - 1) - c\ln\frac{1}{\beta} > 0.$$

We now take $q \in \mathcal{P}_1 \setminus \mathcal{P}_0$. Let, again, $h = \frac{\mathrm{d}\bar{\alpha}}{\mathrm{d}q}$, $g = 1/h$, $\beta_1 = \int_{\mathcal{X}} g\mathrm{d}\bar{\alpha}$, $\theta' = \langle\phi, q\rangle \geq \langle\phi, g\rangle + \phi_o(1 - \beta_1)$ and $\theta_o = \phi_o + c$. The inequality (6.15) now reads

$$F(q^*) - F(q) = -\theta_o + c\int_{\mathcal{X}} \ln f \mathrm{d}\bar{\alpha} + \theta' - c\int_{\mathcal{X}} \ln g\mathrm{d}\bar{\alpha} \geq \theta' - \theta_o - c\ln\big(\frac{\langle\phi, g\rangle - \phi_o\beta_1}{c}\big) \tag{6.17}$$

and

$$\ln\big(\frac{\langle\phi, g\rangle - \phi_o\beta_1}{c}\big) \leq \ln\big(\frac{\theta' - \phi_o(1 - \beta_1) - \phi_o\beta_1}{c}\big) = \ln\big(\frac{\theta' - \phi_o}{c}\big) = \ln\big(1 + \frac{\theta' - \theta_o}{c}\big) \leq \frac{\theta' - \theta_o}{c}.$$

The Jensen inequality is an equality when $g = \frac{\beta_1}{\beta}f$. When $\beta = \beta_1$, then $f = g$ and $F(q^*) - F(q) = \theta' - (\phi_o + c)$. Since $q \neq q^*$, it follows that $\theta' > \langle\phi, f\rangle + (1 - \beta)\phi_o = c + \phi_o$ and so $F(q^*) > F(q)$ (analogously as in the proof of Lemma 6.5). Consider now the case $\beta_1 \neq \beta$. Then

$$F(q^*) - F(q) = \theta' - (\phi_o + c) + c(\ln\beta - \ln\beta_1).$$

Since now

$$\theta' \geq \beta_1/\beta\langle\phi, f\rangle + \phi_o(1 - \beta_1) = \beta_1/\beta(c + \beta\phi_o) + \phi_o(1 - \beta_1) = \frac{\beta_1}{\beta}c + \phi_o,$$

we obtain that $F(q^*) - F(q) \geq \frac{\beta_1}{\beta}c + \phi_o - (\phi_o + c) - c\ln(\frac{\beta_1}{\beta}) = c(\frac{\beta_1}{\beta} - 1) - c\ln(\frac{\beta_1}{\beta}) > 0.$ □

6.2.5. *Proof of the main theorem when $\lambda \in (0, 1)$.*

Proof of Theorem 5.7:
(1) The proof $\bar{\pi}_n \Rightarrow \delta_{q^*}$ is exactly as in Theorem 5.3, just instead of Theorem 5.2, Theorem 5.6 should be used.
(2) Note that $w_n(x) \to 1$ and $\sup_x |w_n(x) - 1| = 1 - \exp[-\frac{\phi_o}{n^\lambda}] \to 0$, since ϕ is continuous ϕ and $\mathcal{X}$ compact. All assumptions of Corollary 3.2 are fulfilled with $r_q = q$. By Theorem 3.1, the limit process $X_1, X_2, \ldots$, exists, and it is an i.i.d. process, where $X_i \sim q^*$.
(3) By Theorem 3.4, $Q_n \Rightarrow \delta_{q^*}$

□

6.2.6. *Proof of the main theorem when $\lambda \geq 1$.*

Proof of Theorem 5.9:
(1) The proof is similar to the proof of Theorem 4.2(2). Since π_n is the $DP(n^{1-\lambda}, \bar{\alpha})$, according to Ghosal and van der Vaart (2017, Theorem 4.16), we have that $\pi_n \Rightarrow \delta_X$ where $X \sim \bar{\alpha}$; let us denote the weak limit by π. To overcome the fact that here $\pi_n \Rightarrow \pi$ instead of $\pi_n = \pi$, we make use of Lemma 4.5: this proves that $\bar{\pi}_n \Rightarrow \pi$.
(2) It is not difficult to prove that $P_n(A_1 \times \cdots \times A_n) = \bar{\alpha}(A_1 \cap \cdots \cap A_n)$ which is the law of the n-dimensional random vector $(X, \ldots, X)$ where $X \sim \bar{\alpha}$.
(3) It is the same as in Theorem 4.2(2).

□

Acknowledgements

The authors wish to thank an anonymous referee for their suggestions, which helped improve clarity of the paper, and Dario Gasbarra for suggesting useful tools for the case of Dirichlet process priors.

References

Aliprantis, C. D. and Border, K. C. *Infinite dimensional analysis: a Hitchhiker's Guide.* Springer, Berlin, third edition (2006). ISBN 978-3-540-32696-0; 3-540-32696-0. MR2378491.

Bansaye, V., Caballero, M.-E., and Méléard, S. Scaling limits of population and evolution processes in random environment. *Electron. J. Probab.*, **24**, Paper No. 19, 38 (2019). MR3925459.

Ben-Ari, I., Matzavinos, A., and Roitershtein, A. On a species survival model. *Electron. Commun. Probab.*, **16**, 226–233 (2011). MR2788894.

Bertacchi, D., Lember, J., and Zucca, F. A stochastic model for the evolution of species with random fitness. *Electron. Commun. Probab.*, **23**, Paper No. 88, 13 (2018). MR3882229.

Bertacchi, D. and Zucca, F. Uniform asymptotic estimates of transition probabilities on combs. *J. Aust. Math. Soc.*, **75** (3), 325–353 (2003). MR2015321.

Bertacchi, D., Zucca, F., and Ambrosini, R. The timing of life history events in the presence of soft disturbances. *J. Theoret. Biol.*, **389**, 287–303 (2016). MR3430974.

Berzunza, G., Sturm, A., and Winter, A. Trait-dependent branching particle systems with competition and multiple offspring. *Electron. J. Probab.*, **26**, Paper No. 153, 41 (2021). MR4347378.

Billingsley, P. *Convergence of probability measures.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition (1999). ISBN 0-471-19745-9. MR1700749.

Dembo, A. and Zeitouni, O. *Large deviations techniques and applications*, volume 38 of *Stochastic Modelling and Applied Probability.* Springer-Verlag, Berlin (2010). ISBN 978-3-642-03310-0. Corrected reprint of the second (1998) edition. MR2571413.

Durrett, R. and Limic, V. Rigorous results for the NK model. *Ann. Probab.*, **31** (4), 1713–1753 (2003). MR2016598.

Durrett, R. and Mayberry, J. Traveling waves of selective sweeps. *Ann. Appl. Probab.*, **21** (2), 699–744 (2011). MR2807971.

Evans, S. N. and Steinsaltz, D. Estimating some features of NK fitness landscapes. *Ann. Appl. Probab.*, **12** (4), 1299–1321 (2002). MR1936594.

Ewens, W. J. *Mathematical population genetics. I. Theoretical introduction*, volume 27 of *Interdisciplinary Applied Mathematics.* Springer, second edition (2004). ISBN 978-1-4419-1898-7. DOI: 10.1007/978-0-387-21822-9.

Feng, S. *The Poisson-Dirichlet distribution and related topics. Models and asymptotic behaviors.* Probability and its Applications (New York). Springer, Heidelberg (2010). ISBN 978-3-642-11193-8. MR2663265.

Fortini, S., Ladelli, L., and Regazzini, E. Exchangeability, predictive distributions and parametric models. *Sankhyā Ser. A*, **62** (1), 86–109 (2000). MR1769738.

Ghosal, S. and van der Vaart, A. *Fundamentals of nonparametric Bayesian inference*, volume 44 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge (2017). ISBN 978-0-521-87826-5. MR3587782.

Guiol, H., Machado, F. P., and Schinazi, R. B. A stochastic model of evolution. *Markov Process. Related Fields*, **17** (2), 253–258 (2011). MR2856242.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G., editors. *Bayesian nonparametrics*, volume 28 of *Cambridge Series in Statistical and Probabilistic Mathematics.* Cambridge University Press, Cambridge (2010). ISBN 978-0-521-51346-3. MR2722987.

Iwasa, Y. and Levin, S. A. The timing of life history events. *J. Theoret. Biol.*, **172** (1), 33–42 (1995). DOI: 10.1006/jtbi.1995.0003.

Lember, J. and Watkins, C. An evolutionary model that satisfies detailed balance. *Methodol. Comput. Appl. Probab.*, **24** (1), 1–37 (2022). MR4379479.

Pfanzagl, J. and Pierlo, W. *Compact systems of sets.* Springer-Verlag, Berlin-New York (1966). MR0216529.

Schweinsberg, J. Rigorous results for a population model with selection I: evolution of the fitness distribution. *Electron. J. Probab.*, **22**, Paper No. 37, 94 (2017a). MR3646063.

Schweinsberg, J. Rigorous results for a population model with selection II: genealogy of the population. *Electron. J. Probab.*, **22**, Paper No. 38, 54 (2017b). MR3646064.