A Revisit of Infinite Population Models for Evolutionary Algorithms on Continuous Optimization Problems

Bo Song   bsong@connect.hku.hk
Department of Electrical and Electronic Engineering, the University of Hong Kong, Pokfulam, Hong Kong

Victor O.K. Li   vli@eee.hku.hk
Department of Electrical and Electronic Engineering, the University of Hong Kong, Pokfulam, Hong Kong

Abstract

Infinite population models are important tools for studying population dynamics of evolutionary algorithms. They describe how the distributions of populations change between consecutive generations. In general, infinite population models are derived from Markov chains by exploiting symmetries between individuals in the population and analyzing the limit as the population size goes to infinity. In this article, we study the theoretical foundations of infinite population models of evolutionary algorithms on continuous optimization problems. First, we show that the convergence proofs in a widely cited study were in fact problematic and incomplete. We further show that the modeling assumption of exchangeability of individuals cannot yield the transition equation. Then, in order to analyze infinite population models, we build an analytical framework based on convergence in distribution of random elements which take values in the metric space of infinite sequences. The framework is concise and mathematically rigorous. It also provides an infrastructure for studying the convergence of the stacking of operators and of iterating the algorithm which previous studies failed to address. Finally, we use the framework to prove the convergence of infinite population models for the mutation operator and the k-ary recombination operator. We show that these operators can provide accurate predictions for real population dynamics as the population size goes to infinity, provided that the initial population is identically and independently distributed.

1   Introduction

Evolutionary algorithms (EAs) are general purpose optimization algorithms with great successes in real-world applications. They are inspired by the evolutionary process in nature. A certain number of candidate solutions to the problem at hand are modeled as individuals in a population. The algorithm evolves the population by mutation, crossover, and natural selection so that individuals with more preferable objective function values have higher survival probability. By the "survival of the fittest" principle, it

is likely that after many generations the population will contain individuals with high fitness values such that they are satisfactory solutions to the problem at hand.

Though conceptually simple, the underlying evolutionary processes and the behaviors of EAs remain to be fully understood. The difficulties lie in the fact that EAs are customizable population-based iterative stochastic algorithms, and the objective function also has great influence on their behaviors. A successful model of EAs should describe both the mechanisms of the algorithm and the influence from the objective function. One way to study EAs is to model them as dynamical systems. The idea is to pick a certain quantity of interest first, such as the distribution of the population or a certain statistic about it. Then, transitions in the state space of the picked quantity are modeled. A transition matrix (when the state space is finite) or a difference equation (when the state space is not finite) for Markov chain is derived to describe how the picked quantity changes between consecutive generations.

In order to characterize the population dynamics accurately, the state space of the Markov chain tends to grow rapidly as the population size increases. As a result, even for time-homogeneous EAs with moderate population size, the Markov chain is often too large to be analyzed or simulated. To overcome this issue, some researchers turn to studying the limiting behaviors of EAs as the population size goes to infinity. The idea is to exploit some kind of symmetry in the state space (such as all individuals having the same marginal distribution), and prove that in the limit the Markov chain can be described by a more compact model. Models built in this way are called infinite population models (IPMs).

In this article, we follow this line of research and study IPMs of EAs on *continuous* space. More specifically, we aim at rigorously proving the convergence of IPMs. In this study, by "convergence" we usually mean that IPMs characterize limiting behaviors of real EAs. "An IPM converges loosely" means that as the population size goes to infinity, the population dynamics of the real EA converge in a sense to the population dynamics predicted by this model. This usage is different from conventional ones where it means that the EA eventually locates and gets stuck in some local or global optima. Convergence results are the foundations and justifications of IPMs.

The main results of the article can be summarized as follows. First, we show that a widely cited research on convergence of IPM was problematic. It is mainly because the core assumption of exchangeability of individuals in their proof cannot lead to the convergence conclusion. Then, we build an analytical framework from a different perspective and show that it defines convergence in general settings. Then, to show the effectiveness of our framework, we prove the convergence of IPM of simple EA with mutation and crossover operators when the initial population follows an identical and independent distribution (i.i.d.). Finally, we discuss the results and point out that the convergence of IPM of simple EA with proportionate selection is yet to be developed.

To our knowledge, there are very few research efforts which directly studied the convergence of IPMs. Among them, Vose (1999b) and Qi and Palmieri (1994a,b) are the classic ones. We focus on Qi and Palmieri (1994a,b) in this article as their results are for EA on continuous solution space and most relevant here. In the first part of their study, the authors built a model of simple EA with only mutation and proportionate selection. A transition equation is constructed which describes how the probability density functions (p.d.f.s) of *marginal* distribution of the population change between consecutive generations. The authors proved that if individuals are exchangeable in a population, as the population size goes to infinity, the marginal p.d.f.s of populations of simple EA

will converge *point-wise* to the p.d.f.s. calculated by the following transition equation:

$$f_{\mathbf{x}_{k+1}}(x) = \frac{\int_{\mathbb{F}} f_{\mathbf{x}_k}(y)g(y)f_w(x|y)\mathrm{d}y}{\int_{\mathbb{F}} f_{\mathbf{x}_k}(y)g(y)\mathrm{d}y}, \tag{1}$$

where $\mathbb{F}$ is the solution space, $f_{\mathbf{x}_k}$ is the predicted marginal p.d.f. of the kth generation, g is the objective function to be maximized, and $f_w(x|y)$ is the conditional p.d.f. decided by the mutation operator. In the second part of the research, the authors further analyzed the crossover operator and modified the transition equation to include all three operators in the simple EA.

In Section 2, we will examine Qi's model and show that their study is *unsound* and *incomplete*. First, we provide a counterexample to show that in the authors' proof a key assertion about the law of large numbers (LLN) for exchangeable random vectors is generally not true. Therefore, the whole proof is unsound. Furthermore, we show that the modeling assumption of exchangeability of individuals cannot yield the transition equation in general. This means that under the authors' modeling assumption, the conclusion (1) cannot be reached.

In addition, we show that the authors' proofs in Qi and Palmieri (1994a,b) are incomplete. The authors did not address the convergence of the stacking of operators and of recursively iterating the algorithm. In essence, the authors attempted to prove the convergence of the IPM for only *one* iteration step. Even if the proof for (1) is correct, it only shows that as $n \to \infty$ the marginal p.d.f.s of the $(k + 1)$th population converges point-wise to $f_{\mathbf{x}_{k+1}}(x)$, provided that the marginal p.d.f. of the kth generation is $f_{\mathbf{x}_k}(x)$. However, this convergence does not automatically hold for all subsequent generations. As a result, (1) cannot be iterated to make predictions for subsequent ($> k + 1$) generations.

Besides Qi and Palmieri (1994a,b), we found no other studies that attempted to prove the convergence of IPMs for EAs on *continuous* space. Therefore, in Section 3 we propose a general analytical framework. The novelty of our framework is that from the very start of the analysis, we model generations of the population as random elements taking values in the metric space of infinite sequences, and we use convergence in distribution instead of point-wise convergence to define the convergence of IPMs.

To illustrate the effectiveness of our framework, we perform convergence analysis of IPM of the simple EA in Sections 4 and 5. In Section 4, we adopted a "stronger" modeling assumption that individuals of the same generation in the IPM are identically and independently distributed (i.i.d.), and we gave sufficient conditions under which the IPM is convergent. For general EA, this assumption may seem restricted at first sight, but it turns out to be a reasonable one. In Section 5, we analyze the mutation operator and the k-ary recombination operator. We show that these commonly used operators have the property of producing i.i.d. populations, in the sense that if the initial population is i.i.d., as the population size goes to infinity, in the limit all subsequent generations are also i.i.d. This means that for these operators, the transition equation in the IPM can predict the real population dynamics as the population size goes to infinity. We also show that our results hold even if these operators are stacked together and iterated repeatedly by the algorithm. Finally, in Section 6 we conclude the article and propose future research.

To be complete, regarding Qi and Palmieri (1994a,b), there is a comment by Yong et al. (1998) with a reply published. However, the comment was mainly about the latter part of Qi and Palmieri (1994a), where the authors analyzed the properties of EAs based on the IPM. It did not discuss the proof for the model itself. For IPMs of EAs on *discrete*

optimization problems, extensive research was done by Vose et al. in a series of studies (Nix and Vose, 1992; Vose, 1999b, 1999a, 2004). The problems under consideration were discrete optimization problems with *finite* solution space. The starting point of the authors' analysis was to model each generation of the population as an "incidence vector," which describes for each point in the solution space the proportion of the population it occupies. Based on this representation the authors derived transition equations between incidence vectors of consecutive generations and analyzed their properties as the population size goes to infinity. However, for EAs on *continuous* solution space, the analyses of Vose et al. are not immediately applicable. This is because for continuous optimization problems the solution space is not denumerable. Therefore, the population cannot be described by a finite-dimensional incidence vector.

2 Discussion of the Works of Qi et al.

In this section we analyze the results of Qi and Palmieri (1994a,b). We begin by introducing some preliminaries for the analysis. Then, in Section 2.2, following the notations and derivations in the authors' papers, we provide a counterexample to show that the convergence proof for the transition equation in Qi and Palmieri (1994a) is problematic. We further show that the modeling assumption of exchangeability cannot yield the transition equation in general. In Section 2.3, we show that the analyses in Qi and Palmieri (1994a,b) are incomplete. The authors did not prove the convergence of IPMs in the cases where operators are stacked together and the algorithm is iterated for multiple generations.

2.1 Preliminaries

In the authors' paper (Qi and Palmieri, 1994a), the problem to be optimized is

$$\arg \max_x g(x) \text{ s.t. } x \in \mathbb{F} \subseteq \mathbb{R}^m, \tag{2}$$

where $\mathbb{F}$ is the solution space and g is some given objective function. The analysis intends to be general; therefore, no explicit form of g is assumed. The algorithm to be analyzed is the simple EA with proportionate selection and mutation. Let $\mathbf{X}_k = (\mathbf{x}_k^j)_{j=1}^N$ denote the kth generation produced by the EA, where N is the population size. To generate the $(k+1)$th population, proportionate selection produces intermediate populations $\mathbf{X}_k'$ following the conditional probability that

$$P(\mathbf{x}_k'^i = \mathbf{x}_k^j | \mathbf{X}_k) = \frac{g(\mathbf{x}_k^j)}{\sum_{l=1}^N g(\mathbf{x}_k^l)}, \text{ for all } i, j = 1, 2, \ldots, N. \tag{3}$$

After selection, each individual in $\mathbf{X}_k'$ is mutated to generate individuals in $\mathbf{X}_{k+1}$. Mutation follows the conditional probability that

$$f(\mathbf{x}_{k+1}^i = x | \mathbf{x}_k'^i = y) = f_w(x|y). \tag{4}$$

Overall the algorithm is illustrated in Figure 1.

After presenting the optimization problem and the algorithm, the authors proved the convergence of the IPM if the distributions of individuals in the population are exchangeable. It is the main result in Qi and Palmieri (1994a).

THEOREM 1 (Theorem 1 in Qi and Palmieri, 1994a): *Assume that the fitness function $g(x)$ in (2) and the mutation operator of simple EA described by (4) satisfy the following conditions:*

Require: population size N; p.d.f. of the initial population $f_{\mathbf{x}_0}$

1: $k \leftarrow 0$

2: sample the N i.i.d. individuals $\mathbf{x}_k^1, \mathbf{x}_k^2, \ldots, \mathbf{x}_k^N$ according to $f_{\mathbf{x}_0}$

3: **while** stopping criteria is not satisfied **do**

4:   **Selection:** select $\mathbf{x}_k'^1, \mathbf{x}_k'^2, \ldots, \mathbf{x}_k'^N$ from $\mathbf{x}_k^1, \mathbf{x}_k^2, \ldots, \mathbf{x}_k^N$ identically and independently according to the probability that $P(\mathbf{x}_k'^i = \mathbf{x}_k^j | \mathbf{X}_k) = \frac{g(\mathbf{x}_k^j)}{\sum_{l=1}^N g(\mathbf{x}_k^l)}, \forall i, j = 1, 2, \ldots, N$

5:   **Mutation:** perturb $\mathbf{x}_k'^1, \mathbf{x}_k'^2, \ldots, \mathbf{x}_k'^N$ to form the new generation $\mathbf{x}_{k+1}^1, \mathbf{x}_{k+1}^2, \ldots, \mathbf{x}_{k+1}^N$ according to the common conditional p.d.f. $f(\mathbf{x}_{k+1}^i = x | \mathbf{x}_{k+1}'^i = y) = f_w(x|y), \forall i = 1, 2, \ldots, N$

6:   $k \leftarrow k + 1$

7: **end while**

Figure 1: The pseudocode of the simple EA.

1. $0 < g_{\min} \le g(x) \le g_{\max} < \infty, \forall x \in \mathbb{F}$.

2. $\sup_{x,y \in \mathbb{R}^d} f_w(x|y) \le M < \infty$.

Then as $n \to \infty$, the time history of the simple EA can be described by a sequence of random vectors $(\mathbf{x}_k)_{k=0}^\infty$ with densities

$$f_{\mathbf{x}_{k+1}}(x) = \frac{\int_{\mathbb{F}} f_{\mathbf{x}_k}(y) g(y) f_w(x|y) \, \mathrm{d}y}{\int_{\mathbb{F}} f_{\mathbf{x}_k}(y) g(y) \, \mathrm{d}y}. \tag{5}$$

In Theorem 1, $f_{\mathbf{x}_k}$ is the marginal p.d.f. of the kth generation predicted by the IPM. It should be emphasized that in Qi and Palmieri (1994a,b), the authors proved this theorem under the assumption that simple EA has exchangeable individuals in the population. Though not explicitly stated in the theorem, the assumption of exchangeability is the core assumption in their proof and an integral part of their formulation of the theorem.

For analyses in this article, we use the concept of exchangeability in probability theory. Its definition and some basic facts are listed.

DEFINITION 1 (Exchangeable random variables, Definition 1.1.1 in Taylor et al., 1985): *A finite set of random variables $\{\mathbf{x}_i\}_{i=1}^n$ is said to be exchangeable if the joint distribution of $(\mathbf{x}_i)_{i=1}^n$ is invariant with respect to permutations of the indices $1, 2, \ldots, n$. A collection of random variables $\{\mathbf{x}_\alpha : \alpha \in \Gamma\}$ is said to be exchangeable if every finite subset of $\{\mathbf{x}_\alpha : \alpha \in \Gamma\}$ is exchangeable.*

Definition 1 can also be extended to cover exchangeable random vectors or exchangeable random elements by replacing the term "random variables" in the definition with the respective term. One property of exchangeability is that if $\{\mathbf{x}_i\}_{i=1}^n$ are n exchangeable random elements, then the joint distributions of any $1 \le k \le n$ distinct ones of them are always the same (Proposition 1.1.1 in Taylor et al., 1985). When $k = 1$ this property indicates that $\{\mathbf{x}_i\}_{i=1}^n$ have the same marginal distribution. Another property is

that a collection of random elements are exchangeable if and only if they are conditionally independent and identically distributed (c.i.i.d.) given some σ-field $\mathcal{G}$ (Theorem 1.2.2 in Taylor et al., 1985). Conversely, a collection of c.i.i.d. random elements are always exchangeable. Finally, it is obvious that i.i.d. random elements are exchangeable, but the converse is not necessarily true.

It can be seen that the simple EA generates c.i.i.d. individuals given the current population. Therefore, the individuals within the same generation are exchangeable, and they have the same marginal distribution. It is the core assumption of the proof in Qi and Palmieri (1994a,b) and Condition 3 in the Theorem 1.

2.2 Convergence Proof of the Transition Equation

In this section, we analyze the proof of Theorem 1 and show that it is incorrect. The proof by Qi et al. is in Appendix A of Qi and Palmieri (1994a). In the proof, the authors assumed that individuals in the same generation are exchangeable; therefore, they have the same marginal distribution. After a series of derivation steps, the authors managed to obtain a transition equation between the density functions of $\mathbf{x}_{k+1}^i$ and $\mathbf{X}_k$:

$$f_{\mathbf{x}_{k+1}^i}(x) = \int \int_{\mathbb{F}^N} \frac{g(y_j) f_w(x|y_j)}{\frac{1}{N} \sum_{l=1}^{N} g(y_l)} f_{\mathbf{X}_k}(y_1, y_2, \ldots, y_n)$$

$$\mathrm{d}y_1 \mathrm{d}y_2 \ldots \mathrm{d}y_n \text{ for any fixed } i, j$$

$$= \mathrm{E}\left[\frac{\boldsymbol{\xi}_k(x)}{\boldsymbol{\eta}_k^N}\right], \tag{6}$$

where in (6),

$$\boldsymbol{\xi}_k(x) \triangleq g(\mathbf{x}_k^j) f_w(x|\mathbf{x}_k^j) \text{ for any fixed } j, \tag{7}$$

$$\boldsymbol{\eta}_k^N \triangleq \frac{1}{N} \sum_{l=1}^{N} g(\mathbf{x}_k^l). \tag{8}$$

Eq. (6) is correct. It accurately describes how the marginal p.d.f. for any individual in the next generation can be calculated from the joint p.d.f. of individuals in the current generation. Noticing that $\boldsymbol{\eta}_k^N$ is the average of the exchangeable random variables $\{g(\mathbf{x}_k^j)\}_{j=1}^{N}$, by the LLN for exchangeable random variables, the authors asserted that

$$\lim_{N \to \infty} \boldsymbol{\eta}_k^N = \boldsymbol{\eta}_k \text{ almost surely (a.s.).} \tag{9}$$

The authors further asserted that $\boldsymbol{\eta}_k$ is itself a random variable, satisfying

$$\mathrm{E}[\boldsymbol{\eta}_k] = \mathrm{E}[g(\mathbf{x}_k^j)] \text{ for any } j. \tag{10}$$

Eqs. (9) and (10) correspond to (A13) and (A14) in Appendix A of Qi and Palmieri (1994a), respectively. The authors' proof is *correct* until this step. However, the authors then asserted that

$\boldsymbol{\eta}_k$ is independent of $\boldsymbol{\eta}_k^N$ for any finite N. In particular, $\boldsymbol{\eta}_k$ is independent of

$$\boldsymbol{\eta}_k^1 = g(\mathbf{x}_k^j) \text{ for all } j = 1, 2, \ldots, N. \tag{11}$$

Based on this assertion the authors then proved that for all k and x,

$$\lim_{N \to 0} \left| \mathrm{E}\left[\frac{\boldsymbol{\xi}_k(x)}{\boldsymbol{\eta}_k^N}\right] - \frac{\mathrm{E}[\boldsymbol{\xi}_k(x)]}{\mathrm{E}[\boldsymbol{\eta}_k]} \right| = 0. \tag{12}$$

As a result, the p.d.f. in (6) converges point-wisely to $\frac{E[\xi_k(x)]}{E[\eta_k]}$, which is equal to the right hand side of (5). Hence the authors claimed that Theorem 1 is proved.

In the following, we provide a counterexample to show that assertion (11) is not true when $N \geq 2$ ($N = 1$ is the degenerate case). Then, we carry out further analysis to show that under the modeling assumption of exchangeability, conclusion (12) or equivalently Theorem 1 cannot be true in general.

2.2.1 On Assertion (11)

We first reformulate the assertion. Since $\{\mathbf{x}_k^l\}_{l=1}^N$ are exchangeable, $\{g(\mathbf{x}_k^l)\}_{l=1}^N$ are exchangeable (Property 1.1.2 in Taylor et al., 1985). Let $\mathbf{y}_l = g(\mathbf{x}_k^l)$, $l = 1, \ldots, N$. Then the premises of Theorem 1 imply

$$\{\mathbf{y}_l\}_{l=1}^N \text{ are exchangeable and } g_{\min} \leq \mathbf{y}_l \leq g_{\max}. \tag{13}$$

Let $\mathbf{y} = \boldsymbol{\eta}_k$. According to (8), (9), and (10), $\mathbf{y}$ has the properties that

$$\begin{cases} \lim_{N \to \infty} \frac{1}{N} \sum_{l=1}^N \mathbf{y}_l = \mathbf{y}, \text{ a.s.,} & (14) \\ E(\mathbf{y}) = E(\mathbf{y}_l) \text{ for any } l. & (15) \end{cases}$$

Since g is a general function, there are no other restrictions for $\{\mathbf{y}_l\}_{l=1}^N$ and $\mathbf{y}$. Therefore, (11) is equivalent to the following assertion:

$$\text{For any } \{\mathbf{y}_l\}_{l=1}^N \text{ and } \mathbf{y} \text{ satisfying (13), (14), and (15), } \mathbf{y} \text{ and } \frac{1}{N} \sum_{l=1}^N \mathbf{y}_l \text{ are}$$

$$\text{independent for any finite } N. \text{ In particular, } \mathbf{y} \text{ is independent of } \mathbf{y}_l \text{ for any } l.$$

$$\tag{16}$$

However, we use the following counterexample (modified from Example 1.1.1 and related discussions on pages 11–12 in Taylor et al. (1985), to show that assertion (16) is not true. Therefore, (11) is not true.

2.2.2 Counterexample

Let $\{\mathbf{z}_l\}_{l=1}^\infty$ be a sequence of i.i.d. random variables satisfying

$$-\frac{g_{\max} - g_{\min}}{4} \leq \mathbf{z}_l \leq \frac{g_{\max} - g_{\min}}{4} \text{ and } E(\mathbf{z}_l) = 0$$

for all l. Let $\mathbf{y}$ be a random variable *independent* of $\{\mathbf{z}_l\}_{l=1}^\infty$ satisfying

$$\frac{g_{\max} + 3g_{\min}}{4} \leq \mathbf{y} \leq \frac{3g_{\max} + g_{\min}}{4}$$

and

$$E(\mathbf{y}) = \frac{g_{\max} + g_{\min}}{2}.$$

Finally, let $\mathbf{y}_l = \mathbf{z}_l + \mathbf{y}$ for all l.

It can easily be verified that $\{\mathbf{y}_l\}_{l=1}^\infty$ and $\mathbf{y}$ satisfy (13) and (15). Since $\mathbf{z}_l$ is bounded, $E(|\mathbf{z}_l|) < \infty$ for any l. By the strong law of large numbers (SLLN) for i.i.d. random variables, $\frac{1}{N} \sum_{l=1}^N \mathbf{z}_l \to 0$ a.s. as $N \to \infty$; therefore, (14) is also satisfied, that is, $\mathbf{y}$ is the limit of $\frac{1}{N} \sum_{l=1}^N \mathbf{y}_l$ as $N \to \infty$. However, because $\frac{1}{N} \sum_{l=1}^N \mathbf{y}_l = \mathbf{y} + \frac{1}{N} \sum_{l=1}^N \mathbf{z}_l$ and $\mathbf{y}$ is independent of $\{\mathbf{z}_l\}_{l=1}^\infty$, it can be seen that $\frac{1}{N} \sum_{l=1}^N \mathbf{y}_l$ is *not* independent of $\mathbf{y}$ except for some degenerate cases (e.g., when $\mathbf{y}$ equals to a constant). In particular, in general $\mathbf{y}_l = \mathbf{y} + \mathbf{z}_l$ is *not* independent of $\mathbf{y}$ for any l. Therefore, assertion (16) is not true. Equivalently, assertion (11) is not true. This renders the authors' proof for (12) invalid.

2.2.3 Further Analysis

In the following, we carry out further analysis to show that (12) cannot be true even considering other methods of proof and adding new sufficient conditions. Therefore, in general, Theorem 1 cannot be true.

To begin with, consider the random variable $\frac{\xi_k(x)}{\eta_k^N}$. We prove the following lemma.

LEMMA 1: $E\left(\frac{\xi_k(x)}{\eta_k^N}\right) \to E\left(\frac{\xi_k(x)}{\eta_k}\right)$ as $N \to \infty$.

PROOF: According to (9), $\eta_k^N \overset{\text{a.s.}}{\to} \eta_k$, since $g_{\min} \le \eta_k^N \le g_{\max}$, $0 < g_{\min} \le \eta_k \le g_{\max}$ almost surely.

Since $h(x) = \frac{1}{x}$ is continuous on $(0, \infty)$,

$$h(\eta_k^N) = \frac{1}{\eta_k^N} \overset{\text{a.s.}}{\to} h(\eta_k) = \frac{1}{\eta_k} \text{ (Proposition 47.2 in Port, 1994)}.$$

Then

$$\frac{\xi_k(x)}{\eta_k^N} \overset{\text{a.s.}}{\to} \frac{\xi_k(x)}{\eta_k} \text{ (Proposition 47.4 (ii) in Port, 1994)}.$$

Finally, by the conditions in Theorem 1, $0 \le \frac{\xi_k(x)}{\eta_k^N} \le \frac{Mg_{\max}}{g_{\min}}$. By the Lebesgue's Dominated Convergence Theorem (Proposition 11.30 in Port, 1994), $E\left(\frac{\xi_k(x)}{\eta_k^N}\right) \to E\left(\frac{\xi_k(x)}{\eta_k}\right)$ as $N \to \infty$. $\square$

Now by Lemma 1, (12) is equivalent to

$$E\left(\frac{\xi_k(x)}{\eta_k}\right) = \frac{E[\xi_k(x)]}{E[\eta_k]}. \tag{$\triangle$}$$

Now it is clear that if the only assumption is exchangeability, ($\triangle$) is *not* true even considering other methods of proof. Of course, if (11) is true, $\xi_k(x)$ and η_k are independent, then ($\triangle$) is true. However, as already shown by the counterexample, (11) is not true in general. Therefore, ($\triangle$), and equivalently Theorem 1, are in general not true.

A natural question then arises: Is it possible to introduce some reasonable sufficient conditions such that ($\triangle$) can be proved? One of such conditions frequently used is that $\eta_k = E[g(\mathbf{x}_k^i)]$. However, the following analysis shows that given the modeling assumption of exchangeability, this condition is not true in general. Therefore, it cannot be introduced.

For exchangeable random variables $\{g(\mathbf{x}_k^l)\}_{l=1}^N$, we have

$$V(\eta_k) = \lim_{N \to \infty} V(\eta_k^N) \tag{17}$$

$$= \lim_{N \to \infty} \left\{ E\left[\frac{\sum_{l=1}^N g(\mathbf{x}_k^l)}{N}\right]^2 - \left[E\frac{\sum_{l=1}^N g(\mathbf{x}_k^l)}{N}\right]^2 \right\}$$

$$= \lim_{N \to \infty} \frac{1}{N^2} \left\{ \sum_{l=1}^N V\left[g(\mathbf{x}_k^l)\right] + \sum_{i=1}^N \sum_{j=1, j \neq i}^N C\left[g(\mathbf{x}_k^i), g(\mathbf{x}_k^j)\right] \right\}$$

Table 1: Population dynamics of EA_n under operator H.

k EA_n	0	1	2	...
EA_1	$\mathbf{P}_0^1$	$H_1(\mathbf{P}_0^1)$	$H_1(H_1(\mathbf{P}_0^1))$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
EA_n	$\mathbf{P}_0^n$	$H_n(\mathbf{P}_0^n)$	$H_n(H_n(\mathbf{P}_0^n))$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	
EA_∞	$\mathbf{P}_0^\infty$	$H_\infty(\mathbf{P}_0^\infty)$	$H_\infty(H_\infty(\mathbf{P}_0^\infty))$	...

$$= \lim_{N\to\infty} \left\{ \frac{V\left[g(\mathbf{x}_k^1)\right]}{N} + \frac{N-1}{N} C\left[g(\mathbf{x}_k^1), g(\mathbf{x}_k^2)\right] \right\} \tag{18}$$

$$= C\left[g(\mathbf{x}_k^1), g(\mathbf{x}_k^2)\right], \tag{19}$$

where $V(\mathbf{x})$ is the variance of $\mathbf{x}$ and $C(\mathbf{x}, \mathbf{y})$ is the covariance of $\mathbf{x}$ and $\mathbf{y}$. (17) is by the boundedness of η_k^N and the Lebesgue's Dominated Convergence Theorem, (18) is by the exchangeability of $\{\mathbf{x}_k^j\}_{j=1}^N$, and (19) is by the boundedness of g and pushing N to infinity. Now it is clear that if the only modeling assumption is exchangeability, there is no guarantee that $C\left[g(\mathbf{x}_k^1), g(\mathbf{x}_k^2)\right] = 0$. Therefore, in general η_k^N does not converge to a constant. Thus this condition cannot be introduced as a sufficient condition for ($\triangle$).

2.3 The Issue of the Stacking of Operators and Iterating the Algorithm

In the following, we discuss IPMs from another perspective and show that the proofs in Qi and Palmieri (1994a,b) are incomplete, as they only consider convergence of one iteration. Because discussion in this section relates very closely to our proposed framework, **we will use our notation from here on in our article consistently**, which is different from the previous two sections where we followed strictly the notations of Qi and Palmieri (1994a,b).

Consider an EA with only one operator. Let the operator be denoted by H. When the population size is n, denote this EA by EA_n and the operator it *actually* uses by H_n. Let $\mathbf{P}_k^n = (\mathbf{x}_{k,i}^n)_{i=1}^n$ denote the kth generation produced by EA_n. Then the transition rules between consecutive generations produced by EA_n can be described by $\mathbf{P}_{k+1}^n = H_n(\mathbf{P}_k^n)$. In Table 1, we write down the population dynamics of EA_n. Each row in Table 1 shows the population dynamics produced by EA_n. In the table $\mathbf{P}_k^n$ is expanded as $[H_n]^k(\mathbf{P}_0^n)$. Let EA_∞ denote the IPM, and $\mathbf{P}_k^\infty = [H_\infty]^k(\mathbf{P}_0^\infty)$ denote the populations predicted by EA_∞. Then we can summarize the results in Qi and Palmieri (1994a) in the following way.

Assume that the initial population comes from a *known* sequence of individuals, represented by $\mathbf{P}_0 = (\mathbf{x}_i)_{i=1}^\infty$. For EA_n, its initial population $\mathbf{P}_0^n$ consists of the first n elements of $\mathbf{P}_0$; that is, $\mathbf{P}_0^n = (\mathbf{x}_i)_{i=1}^n$. Let $\mathbf{P}_0^\infty = \mathbf{P}_0$. This setting represents the fact that EA_n and EA_∞ use the same initial population. H_n can be viewed as operators on $\mathbf{P}_0$ which takes only the first n elements to produce the next generation. Then the authors

essentially proved that

$$H_n(\mathbf{P}_0) \xrightarrow{\text{m.p.w.}} H_\infty(\mathbf{P}_0) \text{ as } n \to \infty, \tag{20}$$

where m.p.w. stands for point-wise convergence of marginal p.d.f.s.

However, apart from the fact that this proof is problematic, the authors' proof covers only *one* iteration step, corresponding to the column-wise convergence of the $k = 1$ column in Table 1. The problem is that even if (21) is true, it does not automatically lead to the conclusion that for the arbitrary kth step, $[H_n]^k(\mathbf{P}_0) \xrightarrow{\text{m.p.w.}} [H_\infty]^k(\mathbf{P}_0)$ as $n \to \infty$. In other words, one has to study whether the transition equation for one step can be iterated recursively to predict populations after multiple steps. In Table 1, this problem corresponds to whether other columns have similar column-wise convergence property when the convergence of the $k = 1$ column is proved.

To give an example, consider the column of $k = 2$ in Table 1. To prove column-wise convergence, the authors need to prove that given (20),

$$H_n(\mathbf{P}_1^n) \xrightarrow{\text{m.p.w.}} H_\infty(\mathbf{P}_1^\infty), \text{ or equivalently} \tag{21}$$

$$[H_n]^2(\mathbf{P}_0) \xrightarrow{\text{m.p.w.}} [H_\infty]^2(\mathbf{P}_0) \tag{22}$$

as $n \to \infty$. Comparing (20) with (21) and (22), (21) has the same sequence of operators but with a sequence of converging inputs, (22) has the same input but with a sequence of different operators. Therefore, they are not necessarily true even if (20) is proved. A similar problem exists when considering the arbitrary kth generation. We call this problem the issue of iterating the algorithm. As studies in Qi and Palmieri (1994a,b) ignored this issue, we believe their proofs are *incomplete*.

The issue of the stacking of operators is similar. Given some operator H satisfying (20) and some operator G satisfying $G_n(\mathbf{P}_0) \xrightarrow{\text{m.p.w.}} G_\infty(\mathbf{P}_0)$ as $n \to \infty$, it is *not* necessarily true that $H_n(G_n(\mathbf{P}_0)) \xrightarrow{\text{m.p.w.}} H_\infty(G_\infty(\mathbf{P}_0))$ as $n \to \infty$. However, the authors in Qi and Palmieri (1994b) totally ignored this issue and combined the transition equations for selection, mutation and crossover together (in Section III of Qi and Palmieri 1994b) without any justification.

In addition, there are several statements in the authors' proofs in Qi and Palmieri (1994b) that are questionable. First, in the first paragraph of Appendix A (the proof for Theorem 1 in that paper), the authors considered a pair of parents $\mathbf{x}_k$ and $\mathbf{x}'_k$ for the uniform crossover operator. $\mathbf{x}_k$ and $\mathbf{x}'_k$ are "drawn from the population independently with the same density of $f_{\mathbf{x}_k} \equiv f_{\mathbf{x}'_k}$." Then, the authors claimed that "the joint density of $\mathbf{x}_k$ and $\mathbf{x}'_k$ is therefore $f_{\mathbf{x}_k} \cdot f_{\mathbf{x}'_k}$." This is simply not true. Two individuals drawn independently from the same population are *conditionally* independent, they are not necessarily independent. In fact, without the i.i.d. assumption, it is very likely that individuals in the same population are dependent. Therefore, the joint density function of $\mathbf{x}_k$ and $\mathbf{x}'_k$ is not necessarily $f_{\mathbf{x}_k} \cdot f_{\mathbf{x}'_k}$, and the authors' proof for Theorem 1 in Qi and Palmieri (1994b) is dubious at best. On the other hand, if the authors' modeling assumption is i.i.d. of individuals for the uniform crossover operator, this assumption is incompatible with the modeling assumption of exchangeability in Qi and Palmieri (1994a) for selection and mutation. Therefore, combining the transition equations for all these three operators is problematic, because the i.i.d. assumption cannot hold beyond one iteration step.

Another issue in Qi and Palmieri (1994b) is that the uniform crossover operator produces two *dependent* offspring at the same time. As a result, after uniform crossover, the intermediate population is not even exchangeable because it has pair-wise dependency between individuals. Then the same incompatible assumption problem arises: that is,

the transition equation for the uniform crossover operator cannot be combined with the transition equations for selection and mutation. Besides, the transition equation for the uniform crossover operator cannot be iterated beyond one step and still hold i.i.d. or exchangeability as its modeling assumption.

In summary, several issues arise from previous studies on IPMs for EAs on continuous optimization problems. Therefore, new frameworks and proof methods are needed for analyzing the convergence of IPMs and studying the issue of the stacking of operators and iterating the algorithm.

3 Proposed Framework

In this section, we present our proposed analytical framework. In constructing the framework we strive to achieve the following three goals.

1. The framework should be general enough to cover real-world operators and to characterize the evolutionary process of real EA.

2. The framework should be able to define the convergence of IPMs and serve as justifications of using them. The definition should match one's intuition and at the same time be mathematically rigorous.

3. The framework should provide an infrastructure to study the issue of the stacking of operators and iterating the algorithm.

The contents of this section roughly reflect the pursuit of the first two goals. The third goal is reflected in the sufficient conditions for convergence and i.i.d. IPM construction in Section 4 and the analyses of the simple EA in Section 5. More specifically, in Section 3.1, we introduce notations and preliminaries for the remainder of this article. In Section 3.2, we present our framework. In the framework, each generation is modeled by a random sequence. This approach unifies the spaces of random elements modeling populations of different sizes. In Section 3.3, we define the convergence of the IPM as convergence in distribution on the space of random sequences. We summarize and discuss our framework in Section 3.4.

3.1 Notations and Preliminaries

In the remainder of this article, we focus on the unconstrained continuous optimization problem

$$\arg\max_{x} g(x) \text{ s.t. } x \in \mathbb{R}^d, \tag{23}$$

where g is some given objective function. Our framework is general enough such that it does not require other conditions on the objective function g. However, to prove the convergence of IPMs for mutation and recombination, conditions such as those in Theorem 1 are sometimes needed. We will introduce them when they are required.

From now on we use $\mathbb{N}$ to denote the set of nonnegative integers and $\mathbb{N}_+$ the set of positive integers. For any two real numbers a and b, let $a \wedge b$ be the smaller one of them and $a \vee b$ be the larger one of them. Let $\mathbf{x}, \mathbf{y}$ be random elements of some measurable space $(\Omega, \mathcal{F})$. We use $\mathcal{L}(\mathbf{x})$ to represent the law of $\mathbf{x}$. If $\mathbf{x}$ and $\mathbf{y}$ follow the same law, that is, $P(\mathbf{x} \in A) = P(\mathbf{y} \in A)$ for every $A \in \mathcal{F}$, we write $\mathcal{L}(\mathbf{x}) = \mathcal{L}(\mathbf{y})$. Note that $\mathcal{L}(\mathbf{x}) = \mathcal{L}(\mathbf{y})$ and $\mathbf{x} = \mathbf{y}$ have different meanings. In particular, $\mathbf{x} = \mathbf{y}$ indicates dependency between $\mathbf{x}$ and $\mathbf{y}$.

We use the notation $(x_i)_{i=m}^n$ to represent the array $(x_m, x_{m+1}, \ldots, x_n)$. When $n = \infty$, $(x_i)_{i=m}^\infty$ represents the infinite sequence $(x_m, x_{m+1}, \ldots)$. $\{x_i\}_{i=m}^n$ and $\{x_i\}_{i=m}^\infty$ represent the

collections $\{x_m, x_{m+1}, \ldots, x_n\}$ and $\{x_i | i = m, m + 1, \ldots\}$, respectively. When the range is clear, we use $(x_i)_i$ and $\{x_i\}_i$ or (x_i) and $\{x_i\}$ for short.

Let $\mathbb{S}$ denote the solution space $\mathbb{R}^d$. This simplifies the notation system when we discuss the spaces $\mathbb{S}^n$ and $\mathbb{S}^\infty$. In the following, we define metrics and σ-fields on $\mathbb{S}$, $\mathbb{S}^n$ and $\mathbb{S}^\infty$ and state properties of the corresponding measurable spaces.

$\mathbb{S}$ is equipped with the ordinary metric $\rho(x, y) = [\sum_{i=1}^{d}(x_i - y_i)^2]^{\frac{1}{2}}$. Let $\mathcal{S}$ denote the Borel σ-field on $\mathbb{S}$ generated by the open sets under ρ. Together $(\mathbb{S}, \mathcal{S})$ defines a measurable space.

Similarly, $\mathbb{S}^n$ is equipped with the metric $\rho_n(x, y) = [\sum_{i=1}^{n} \rho^2(x_i, y_i)]^{\frac{1}{2}}$, and the corresponding Borel σ-field under ρ_n is denoted by $\mathcal{S}'^n$. Together $(\mathbb{S}^n, \mathcal{S}'^n)$ is the measurable space for n tuples.

Next, consider the space of infinite sequences $\mathbb{S}^\infty = \{(x_1, x_2, \ldots) \mid x_i \in \mathbb{S}, i \in \mathbb{N}_+\}$. It is equipped with the metric $\rho_\infty(x, y) = \sum_{i=1}^{\infty} \frac{1}{2^i} \cdot \frac{\rho(x_i, y_i)}{1 + \rho(x_i, y_i)}$. The Borel σ-field on $\mathbb{S}^\infty$ under ρ_∞ is denoted by $\mathcal{S}'^\infty$. Then $(\mathbb{S}^n, \mathcal{S}'^\infty)$ is the measurable space for infinite sequences.

Since $\mathbb{S}$ is separable and complete, it can be proved that $\mathbb{S}^n$ and $\mathbb{S}^\infty$ are also separable and complete (Appendix M6 in Billingsley, 1999). In addition, because of separability, the Borel σ-fields $\mathcal{S}'^n$ and $\mathcal{S}'^\infty$ are equal to $\mathcal{S}^n$ and $\mathcal{S}^\infty$, respectively. In other words, the Borel σ-fields $\mathcal{S}'^n$ and $\mathcal{S}'^\infty$ generated by the collection of open sets under the corresponding metrics coincide with the product σ-fields generated by all measurable rectangles ($\mathcal{S}^n$) and all measurable cylinder sets ($\mathcal{S}^\infty$), respectively (Lemma 1.2 in Kallenberg, 2002). Therefore, from now on we write $\mathcal{S}^n$ and $\mathcal{S}^\infty$ for the corresponding Borel σ-fields. Finally, let $\mathbb{M}$, $\mathbb{M}^n$ and $\mathbb{M}^\infty$ denote the set of all random elements of $\mathbb{S}$, $\mathbb{S}^n$ and $\mathbb{S}^\infty$, respectively.

Let $\pi_n : \mathbb{S}^\infty \to \mathbb{S}^n$ be the natural projection: $\pi_n(x) = (x_1, x_2, \ldots, x_n)$. Since given $\mathbf{x} \in \mathbb{M}^\infty$, $(\pi_n \circ \mathbf{x}) : \Omega \to \mathbb{S}^n$ defines a random element of $\mathbb{S}^n$ projected from $\mathbb{S}^\infty$, we also use π_n to denote the mapping: $\pi_n : \mathbb{M}^\infty \to \mathbb{M}^n$ where $\pi_n(\mathbf{x}) = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. By definition, π_n is the operator which truncates random sequences to random vectors. Given $A \subset \mathbb{S}^\infty$, we use $\pi_n(A)$ to denote the projection of A; that is, $\pi_n(A) = \{x \in \mathbb{S}^n : x = \pi_n(y) \text{ for some } y \in A\}$.

3.2 Analytical Framework for EA and IPMs

In this section, we present an analytical framework for the EA and IPMs. First, the modeling assumptions are stated. We deal only with operators that generate c.i.i.d. individuals. Then, we present an abstraction of the EA and IPMs. This abstraction serves as the basis for building our framework. Finally, the framework is presented. It unifies the range spaces of the random elements and defines the convergence of IPMs.

3.2.1 Modeling Assumptions

We assume that the EA on the problem (23) is time homogeneous and Markovian, such that the next generation depends only on the current one, and the transition rule from the kth generation to the $(k + 1)$th generation is invariant with respect to $k \in \mathbb{N}$. We further assume that individuals in the next generation are c.i.i.d. given the current generation. As this assumption is the only extra assumption introduced in the framework, it may need some further explanation.

The main reason for introducing this assumption is to simplify the analysis. Conditional independence implies exchangeability, therefore individuals in the same generation $k \in \mathbb{N}_+$ are always exchangeable. As a result, it is possible to exploit the symmetry in the population and study the transition equations of marginal distributions. Besides, it is because of conditional independence that we can easily expand the random

elements modeling finite-sized populations to random sequences, and therefore define convergence in distribution for random elements of the corresponding metric space. In addition, many real world operators in EAs satisfy this assumption, such as the proportionate selection operator and the crossover operator analyzed in Qi and Palmieri (1994a,b). Finally, we emphasize that exchangeability is a property that facilitates the analysis of convergence across multiple iterations and different operators. Because of the generational nature of EA, we want a property that leads to convergence for one iteration and also holds as a premise for the analysis of the next iteration. Exchangeability serves as this property in our analysis. By exchangeability the convergence results can be extended to further iterations and stacked with results from other operators.

However, we admit that there are some exceptions to our assumption. A most notable one may be the mutation operator, though it does not pose significant difficulties. The mutation operator perturbs each individual in the current population independently, according to a common conditional p.d.f. If the current population is not exchangeable, then after mutation the resultant population is not exchangeable, either. Therefore, it seems that mutation does not produce c.i.i.d. individuals. However, considering the fact that mutation is often used along with other operators, as long as these other operators generate c.i.i.d. populations, the individuals after mutation will be c.i.i.d., too. Therefore, a combined operator of mutation and any other operator satisfying the c.i.i.d. assumption can satisfy our assumption. An example can be seen in Qi and Palmieri (1994a), where mutation is analyzed together with proportionate selection. On the other hand, an algorithm which only uses mutation is very simple. It can be readily modeled and analyzed without much difficulty.

Perhaps more significant exceptions are operators such as selection *without* replacement, or the crossover operator which produces two dependent offspring at the same time. In fact, for these operators not satisfying the c.i.i.d. assumption, it is still possible to expand the random elements modeling finite-sized population to random sequences. For example, the random elements can be padded with some fixed constants or random elements of known distributions to form the random sequences. In this way, our definition of the convergence of IPMs can still be applied. However, whether in this scenario convergence in distribution for these random sequences can still yield meaningful results similar to the transition equation is another research problem. It may need further investigation. Nonetheless, our assumption is equivalent to the exchangeability assumption generally used in previous studies.

3.2.2 The Abstraction of EA and IPMs

Given the modeling assumptions, we develop an abstraction to describe the population dynamics of the EA and IPMs.

Let the EA with population size n be denoted by EA_n, and the kth ($k \in \mathbb{N}$) generation it produces be modeled as a random element $\mathbf{P}_k^n = (\mathbf{x}_{k,i}^n)_{i=1}^n \in \mathbb{M}^n$, where $\mathbf{x}_{k,i}^n \in \mathbb{M}$ is a random element representing the ith individual in $\mathbf{P}_k^n$. Without loss of generality, assume that the EA has two operators, G and H. In each iteration, the EA first employs G on the current population to generate an intermediate population, on which it then employs H to generate the next population. Notice that here G and H are just terms representing the operators in the real EA. They facilitate describing the evolutionary process. For EA_n, G and H are actually *instantiated* as functions from $\mathbb{M}^n$ to $\mathbb{M}^n$, denoted by G_n and H_n, respectively. For example, if G represents proportionate selection, the function $G_n : \mathbb{M}^n \to \mathbb{M}^n$ is the actual operator in EA_n generating n c.i.i.d. individuals according to the conditional probability (3). Of course, for the above abstraction to

be valid, the operators used in EA_n should actually produce random elements in $\mathbb{M}^n$; that is, the newly generated population should be measurable on $(\mathbb{S}^n, \mathcal{S}^n)$. As most operators in real EAs satisfy this condition and this is the assumption implicitly taken in previous studies, we assume that this condition is automatically satisfied.

Given these notations, the evolutionary process of EA_n can be described by the sequence $(\mathbf{P}_k^n)_{k=0}^{\infty}$, where the initial population $\mathbf{P}_0^n$ is known and the generation of $\mathbf{P}_k^n$, $k \in \mathbb{N}_+$ follows the recurrence equation

$$\mathbf{P}_{k+1}^n = (\text{H}_n \circ \text{G}_n)(\mathbf{P}_k^n). \tag{24}$$

Then understanding the population dynamics of the EA can be achieved by studying the distributions and properties of $\mathbf{P}_k^n$.

Let the IPM of the EA be denoted by EA_∞. The population dynamics it produces can be described by the sequence $(\mathbf{P}_k^\infty \in \mathbb{M}^\infty)_{k=0}^{\infty}$, where $\mathbf{P}_0^\infty$ is known and the generation of $\mathbf{P}_k^\infty$, $k \in \mathbb{N}_+$ follows the recurrence equation

$$\mathbf{P}_{k+1}^\infty = (\text{H}_\infty \circ \text{G}_\infty)(\mathbf{P}_k^\infty), \tag{25}$$

in which $\text{G}_\infty, \text{H}_\infty : \mathbb{M}^\infty \to \mathbb{M}^\infty$ are operators in EA_∞ modeled after G and H. Then, the convergence of EA_∞ basically requires that $(\mathbf{P}_k^n)_{n=1}^{\infty}$ converges to $\mathbf{P}_k^\infty$ for every generation k.

3.2.3 The Proposed Framework

As stated before, for each generation $k \in \mathbb{N}$, the elements of the sequence $(\mathbf{P}_k^1, \mathbf{P}_k^2, \ldots)$ and the limit $\mathbf{P}_k^\infty$ are all random elements of different metric spaces. Therefore, the core of developing our model is to expand $\mathbf{P}_k^n$ to random sequences, while ensuring that this expansion will not affect modeling the evolutionary process of the real EA. The result of this step is the sequence of random sequences $(\mathbf{Q}_k^n \in \mathbb{M}^\infty)_{k=0}^{\infty}$ for each $n \in \mathbb{N}_+$, which completely describes the population dynamics of EA_n. For the population dynamics of EA_∞, we just let $\mathbf{Q}_k^\infty = \mathbf{P}_k^\infty$.

The expansion of $\mathbf{P}_k^n$ and the relationships between $\mathbf{P}_k^n$, $\mathbf{Q}_k^n$, and $\mathbf{Q}_k^\infty$ are the core of our framework. In the following, we present them rigorously.

3.2.4 The Expansion of $\mathbf{P}_k^n$

We start by decomposing each of G_n and H_n to two operators. One operator is from $\mathbb{S}^\infty$ to $\mathbb{S}^n$. It corresponds to how to convert random sequences to random vectors. A natural choice is the projection operator π_n.

To model the evolutionary process, we also have to define how to expand random vectors to random sequences. In other words, we have to define the expansions of G_n and H_n, which are functions from $\mathbb{S}^n$ to $\mathbb{S}^\infty$.

DEFINITION 2 (The expansion of operator): *For an operator* $\text{T}_n : \mathbb{M}^n \to \mathbb{M}^n$ *satisfying the condition that for any* $\mathbf{x} \in \mathbb{M}^n$, *the elements of* $\text{T}_n(\mathbf{x})$ *are c.i.i.d. given* $\mathbf{x}$, *the expansion of* T_n *is the operator* $\widetilde{\text{T}}_n : \mathbb{M}^n \to \mathbb{M}^\infty$, *satisfying that for any* $\mathbf{x} \in \mathbb{M}^n$,

1. $\text{T}_n(\mathbf{x}) = (\pi_n \circ \widetilde{\text{T}}_n)(\mathbf{x})$.

2. *The elements of* $\widetilde{\text{T}}_n(\mathbf{x})$ *are c.i.i.d. given* $\mathbf{x}$.

In Definition 2, the operator $\widetilde{\text{T}}_n$ is the expansion of T_n. Condition 1 ensures that T_n can be safely replaced by $\pi_n \circ \widetilde{\text{T}}_n$. Condition 2 ensures that the paddings for the sequence are generated according to the same conditional probability distribution as that used by T_n to generate new individuals. In other words, if the operator $\dot{\text{T}}_n : \mathbb{M}^n \to \mathbb{M}$ describes how T_n generates each new individual from the current population, T_n is equivalent to
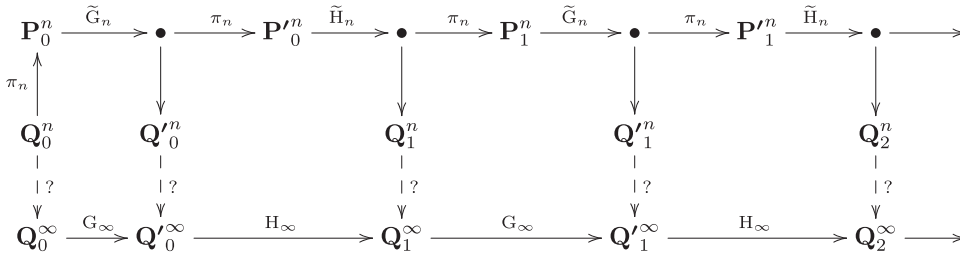
Figure 2: Relationships between $\mathbf{P}_k^n$, $\mathbf{Q}_k^n$, and $\mathbf{Q}_k^\infty$.

invoking $\dot{\mathrm{T}}_n$ independently on the current population for n times, and $\widetilde{\mathrm{T}}_n$ is equivalent to invoking $\dot{\mathrm{T}}_n$ independently for infinite times. Finally, because T_n satisfies the condition in the premise, the expansion $\widetilde{\mathrm{T}}_n$ always exists.

By Definition 2, the operators in EA_n can be decomposed as $\mathrm{G}_n = \pi_n \circ \widetilde{\mathrm{G}}_n$ and $\mathrm{H}_n = \pi_n \circ \widetilde{\mathrm{H}}_n$, respectively. Then, the evolutionary process of EA_n can be described by the sequence of random sequences $[\mathbf{Q}_k^n = (\mathbf{y}_{k,i}^n)_{i=0}^\infty \in \mathbb{M}^\infty]_{k=0}^\infty$, satisfying the recurrence equation

$$\mathbf{Q}_{k+1}^n = (\widetilde{\mathrm{H}}_n \circ \pi_n \circ \widetilde{\mathrm{G}}_n)(\mathbf{P}_k^n), \tag{26}$$

where $\mathbf{P}_k^n$ follows the recurrence equation (24), and $\mathbf{Q}_0^n = (\mathbf{P}_0^n, 0, 0, \ldots)$. It can also be proved that

$$\mathbf{P}_k^n = \pi_n(\mathbf{Q}_k^n). \tag{27}$$

Essentially, (26) and (27) describe how the algorithm progresses in the order $\ldots, \mathbf{Q}_k^n, \mathbf{P}_k^n, \mathbf{Q}_{k+1}^n, \mathbf{P}_{k+1}^n, \ldots$. It fully characterizes the population dynamics $(\mathbf{P}_k^n)_k$, and it is clear that the extra step of generating $\mathbf{Q}_k^n$ does not introduce modeling errors.

For EA_∞, because $\mathbf{P}_k^\infty \in \mathbb{M}^\infty$, there is no need for expansion. For convenience we simply let

$$\mathbf{Q}_k^\infty = \mathbf{P}_k^\infty \tag{28}$$

for $k \in \mathbb{N}$.

In summary, the relationships between $\mathbf{P}_k^n$, $\mathbf{Q}_k^n$ and $\mathbf{Q}_k^\infty$ are better illustrated in Figure 2. This is the core of our framework for modeling the EA and IPMs. For clarity, we also show the intermediate populations generated by G (denoted by $\mathbf{P}_k''^n$), their expansions (denoted by $\mathbf{Q}_k''^n$), and their counterparts generated by G_∞ (denoted by $\mathbf{Q}_k'^\infty$), respectively. How they fit in the evolutionary process can be clearly seen in the figure.

In Figure 2, a solid arrow with an operator on it means that the item at the arrow head equals the result of applying the operator on the item at the arrow tail. For example, from the figure it can be read that $\mathbf{Q}_1^n = \widetilde{\mathrm{H}}_n(\mathbf{P}_0''^n)$. Dashed arrow with a question mark on it signals the place to check whether convergence in distribution holds. For example, when $k = 2$, it should be checked whether $(\mathbf{Q}_2^n)_{n=1}^\infty$ converges to $\mathbf{Q}_2^\infty$ as $n \to \infty$.

Finally, one distinction needs special notice. For EA_m and EA_n ($m \neq n$), consider the operators to generate $\mathbf{P}_k^m$ and $\mathbf{P}_k^n$. It is clear that $\mathrm{G}_m : \mathbb{M}^m \to \mathbb{M}^m$ and $\mathrm{G}_n : \mathbb{M}^n \to \mathbb{M}^n$ are two different operators because their domains and ranges are all different. The distinction still exists when we consider $\mathbf{Q}_k^n$, though it is more subtle and likely to be ignored. In Figure 2, if we consider the operator $\widehat{\mathrm{G}}_n = \pi_n \circ \widetilde{\mathrm{G}}_n : \mathbb{M}^\infty \to \mathbb{M}^\infty$, it is clear that $\widehat{\mathrm{G}}_n$ uses the same mechanism to generate new individuals as the one used in $\mathrm{G}_n = \widetilde{\mathrm{G}}_n \circ \pi_n$, and $\mathbf{Q}_k''^n = \widehat{\mathrm{G}}_n(\mathbf{Q}_k^n)$ describes the same population dynamics as that generated by $\mathbf{P}_k''^n = \mathrm{G}_n(\mathbf{P}_k^n)$. However, if we choose $m \neq n$, $\widehat{\mathrm{G}}_m$ and $\widehat{\mathrm{G}}_n$ are both functions from

$\mathbb{M}^\infty$ to $\mathbb{M}^\infty$. Therefore, checking domains and ranges are not enough to discern $\widehat{G}_m$ and $\widehat{G}_n$. It is important to realize that the distinction between $\widehat{G}_m$ and $\widehat{G}_n$ lies in the *contents* of the functions. $\widehat{G}_m$ and $\widehat{G}_n$ use m and n individuals in the current population to generate the new population, respectively, although the new population contains infinite number of individuals. In short, EA_m and EA_n are the EA instantiated with different population sizes. Mathematically, the corresponding population dynamics are modeled by stochastic processes involving *different* operators, even though their domains and ranges may be the same. The same conclusion also holds for the operator H.

3.3 Convergence of IPMs

Given the framework modeling the EA and IPMs, first, we define convergence in distribution for random elements of $\mathbb{S}^\infty$. This is standard material. Then, the convergence of IPMs is defined by requiring that the sequence $(\mathbf{Q}_k^1, \mathbf{Q}_k^2, \ldots)$ converges to $\mathbf{Q}_k^\infty$ for every $k \in \mathbb{N}$.

3.3.1 Convergence in Distribution

As $\mathbf{Q}_k^n$ are random elements of $\mathbb{S}^\infty$, in the following we define convergence in distribution for sequences of $\mathbb{S}^\infty$-valued random elements. Convergence in distribution is equivalent to weak convergence of *induced* probability measures of the random elements. We use the former theory because when modeling individuals and populations as random elements, the former theory is more intuitive and straightforward. The following materials are standard. They contain the definition of convergence in distribution for random elements, as well as some useful definitions and theorems which are used in our analysis of the simple EA. Most of the materials are collected from the theorems and examples in Sections 1–3 of Billingsley (1999). The definition of Prokhorov metric is collected from Section 11.3 in Dudley (2002).

Let $\mathbf{x}, \mathbf{y}, \mathbf{x}_n, n \in \mathbb{N}_+$ be random elements defined on a hidden probability space $(\Omega, \mathcal{F}, P)$ taking values in some separable metric space $\mathbb{T}$. $\mathbb{T}$ is coupled with the Borel σ-field $\mathcal{T}$. Let $(\mathbb{T}', \mathcal{T}')$ be a separable measurable space other than $(\mathbb{T}, \mathcal{T})$.

DEFINITION 3 (Convergence in distribution): *If the sequence $(\mathbf{x}_n)_{n=1}^\infty$ satisfies the condition that $E[h(\mathbf{x}_n)] \to E[h(\mathbf{x})]$ for every bounded, continuous function $h : \mathbb{T} \to \mathbb{R}$, we say $(\mathbf{x}_n)_{n=1}^\infty$ converges in distribution to $\mathbf{x}$, and write $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.*

For $\epsilon > 0$, let $A^\epsilon = \{y \in \mathbb{T} : d(x, y) < \epsilon \text{ for some } x \in A\}$. Then it is well known that convergence in distribution on separable metric spaces can be metricized by the Prokhorov metric.

DEFINITION 4 (Prokhorov metric): *For two random elements $\mathbf{x}$ and $\mathbf{y}$, the Prokhorov metric is defined as*

$$\rho_d(\mathbf{x}, \mathbf{y}) = \inf\{\epsilon > 0 : P(\mathbf{x} \in A) \leq P(\mathbf{y} \in A^\epsilon) + \epsilon, \forall A \in \mathcal{T}\}.$$

Call a set A in $\mathcal{T}$ an $\mathbf{x}$-continuity set if $P(\mathbf{x} \in \partial A) = 0$, where ∂A is the boundary set of A.

THEOREM 2 (The Portmanteau theorem): *The following statements are equivalent.*

1. $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$.

2. $\limsup_n P(\mathbf{x}_n \in F) \leq P(\mathbf{x} \in F)$ *for all closed set $F \in \mathcal{T}$.*

3. $\liminf_n P(\mathbf{x}_n \in G) \geq P(\mathbf{x} \in G)$ *for all open $G \in \mathcal{T}$.*

4. $P(\mathbf{x}_n \in A) \to P(\mathbf{x} \in A)$ *for all $\mathbf{x}$-continuity set $A \in \mathcal{T}$.*

THEOREM 3 (The mapping theorem): *Suppose $h : (\mathbb{T}, \mathcal{T}) \to (\mathbb{T}', \mathcal{T}')$ is a measurable function. Denote by D_h the set of discontinuities of h. If $\mathbf{x}_n \overset{d}{\to} \mathbf{x}$ and $P(D_h) = 0$, then $h(\mathbf{x}_n) \overset{d}{\to} h(\mathbf{x})$.*

Let $\mathbf{a}$, $\mathbf{a}_n$ be random elements of $\mathbb{T}$, $\mathbf{b}$, $\mathbf{b}_n$ be random elements of $\mathbb{T}'$, then $(\mathbf{a}\ \mathbf{b})^T$ and $(\mathbf{a}_n\ \mathbf{b}_n)^T$ are random elements of $\mathbb{T} \times \mathbb{T}'$. Note that $\mathbb{T} \times \mathbb{T}'$ is separable.

THEOREM 4 (Convergence in distribution for product spaces): *If $\mathbf{a}$ is independent of $\mathbf{b}$ and $\mathbf{a}_n$ is independent of $\mathbf{b}_n$ for all $n \in \mathbb{N}_+$, then $(\mathbf{a}_n\ \mathbf{b}_n)^T \overset{d}{\to} (\mathbf{a}\ \mathbf{b})^T$ if and only if $\mathbf{a}_n \overset{d}{\to} \mathbf{a}$ and $\mathbf{b}_n \overset{d}{\to} \mathbf{b}$.*

Theorem 4 is adapted from Theorem 2.8 (ii) in Billingsley (1999).
Let $\mathbf{z}$, $\mathbf{z}_n$, $n \in \mathbb{N}_+$ be random elements of $\mathbb{S}^\infty$.

THEOREM 5 (Finite-dimensional convergence): *$\mathbf{z}_n \overset{d}{\to} \mathbf{z}$ if and only if $\pi_m(\mathbf{z}_n) \overset{d}{\to} \pi_m(\mathbf{z})$ for any $m \in \mathbb{N}_+$.*

Theorem 5 basically asserts that convergence in distribution for countably infinite dimensional random elements can be studied through their finite-dimensional projections. It is adapted from Example 1.2 and Example 2.4 in Billingsley (1999). In Billingsley (1999), the metric space under consideration is $\mathbb{R}^\infty$. However, as both $\mathbb{R}$ and $\mathbb{S}$ are separable, it is not difficult to adapt the proofs for $\mathbb{R}^\infty$ to a proof for Theorem 5. Note that $\pi_m(\mathbf{z})$ are random elements defined on $(\Omega, \mathcal{F}, P)$ taking values in $(\mathbb{S}^m, \mathcal{S}^m)$, and $P[\pi_m(\mathbf{z}) \in A] = P(\mathbf{z} \in A \times \mathbb{S} \times \mathbb{S} \times \ldots)$ for every $A \in \mathcal{S}^m$. The same is true for $\pi_m(\mathbf{z}_n)$.

3.3.2 Convergence of IPM

As convergence in distribution is properly defined, we can use the theory to define convergence of IPMs. The idea is that IPM is convergent (thus justifiable) if and only if it can predict the limit distribution of the population dynamics of EA_n for *every* generation $k \in \mathbb{N}$ as the population size n goes to infinity. It captures the limiting behaviors of real EAs.

DEFINITION 5 (Convergence of IPMs): *An infinite population model EA_∞ is convergent if and only if for every $k \in \mathbb{N}$, $\mathbf{Q}_k^n \overset{d}{\to} \mathbf{Q}_k^\infty$ as $n \to \infty$, where $\mathbf{Q}_k^n$, $\mathbf{Q}_k^\infty$ and the underling $\mathbf{P}_k^n$, $\mathbf{P}_k^\infty$ are generated according to (26), (28), (24), and (25).*

Definition 5 is essentially the core of our proposed framework. It defines the convergence of IPM and is rigorous and clear.

3.4 Summary

In this section, we built a framework to analyze the convergence of IPMs. The most significant feature of the framework is that we model the populations as random sequences, thereby unifying the ranges of the random elements in a common metric space. Then, we gave a rigorous definition for the convergence of IPMs based on the theory of convergence in distribution.

Our framework is general. It only requires that operators produce c.i.i.d. individuals. In fact, any EA and IPM satisfying this assumption can be put into the framework. However, to obtain meaningful results, the convergence of IPMs has to be proved. This may require extra analyses on IPM and the inner mechanisms of the operators. These analyses are presented in Sections 4 and 5.

Finally, there is one question worth discussing. In our framework, the expansion of operator is carried out by padding the finite population with c.i.i.d. individuals following the *same* marginal distribution. Then a question naturally arises: why not pad the finite population with some other random elements, or just with the constant 0? This

idea deserves consideration. After all, if the expansion is conducted by padding 0s, the requirement of c.i.i.d. can be discarded, and the framework and the convergence of IPMs stay the same. However, we did not choose this approach. The reason is that padding the population with c.i.i.d. individuals facilitates the analysis of the IPM. For example, in our analysis in Sections 4 and 5, the sufficient conditions for the convergence of IPMs require us to consider $\Gamma_m(\mathbf{Q}_k^n)$, where Γ is the operator under analysis. Γ_m uses the first m elements of $\mathbf{Q}_k^n$ to generate new individuals. Now if $m > n$ and $\mathbf{Q}_k^n$ is expanded from $\mathbf{P}_k^n$ by padding 0s, $\Gamma_m(\mathbf{Q}_k^n)$ does not make any sense because the m individuals used by Γ_m have $(m - n)$ 0s. This restricts our option in proving the convergence of IPMs.

4   Sufficient Conditions for Convergence of IPMs and I.I.D. IPM Construction

In this section, we give applicable sufficient conditions for convergence of IPMs. In Section 4.1, we give sufficient conditions for the convergence of IPMs. To appreciate the necessity, consider the framework in Figure 2. To prove the convergence of IPM, by Definition 5, we should check whether $\mathbf{Q}_k^n \xrightarrow{\text{d}} \mathbf{Q}_k^\infty$ as $n \to \infty$ for every $k \in \mathbb{N}$. However, this direct approach is usually not viable. To manually check the convergence for all values of k is wearisome and sometimes difficult. This is because as k increases, the distributions of $\mathbf{Q}_k^n$ and $\mathbf{Q}_k^\infty$ change. Therefore, the method needed to prove $\mathbf{Q}_k^n \xrightarrow{\text{d}} \mathbf{Q}_k^\infty$ as $n \to \infty$ may be different from the method needed to prove $\mathbf{Q}_{k+1}^n \xrightarrow{\text{d}} \mathbf{Q}_{k+1}^\infty$ as $n \to \infty$. Of course, after proving the cases for several values of k, it may be possible to discover some patterns in the proofs, which can be extended to cover other values of k, thus proving the convergence of the IPM. But this process is still tedious and uncertain.

In view of this, a "smarter" way to prove the convergence of IPM may be the following method. First, the convergence of IPM for *one* iteration step for *each* operator is proved. Then, the results are combined and extended to cover the whole population dynamics. The idea is that if the convergence holds for one generation number k, then it can be passed on automatically to all subsequent generations. For example, in Figure 2, consider the operators G_∞ and $\widetilde{G}_n \circ \pi_n$. The first step is to prove that

$$\text{if } \mathbf{Q}_k^n \xrightarrow{\text{d}} \mathbf{Q}_k^\infty \text{ as } n \to \infty, \text{ then } \mathbf{Q}_k'^n \xrightarrow{\text{d}} \mathbf{Q}_k'^\infty \text{ as } n \to \infty. \tag{29}$$

In other words, G_∞ can model $\widetilde{G}_n \circ \pi_n$ for one iteration step. Then, after obtaining similar results for H_∞ and $\widetilde{H}_n \circ \pi_n$, we combine the results together and the convergence of the overall IPM is proved.

However, this approach still seems difficult because we have to prove this pass-on relation (30) holds for every k. In essence, this corresponds to whether the operators in IPM can be stacked together and iterated for any number of steps. This is the issue of the stacking of operators and iterating the algorithm. Therefore, in Section 4.1, we give sufficient conditions for this to hold. These conditions are important. If they hold, proving the convergence of the overall IPM can be broken down to proving the convergence of one iteration step of each operator in IPM. This greatly reduces the difficulty in deriving the proof.

To model real EAs, IPM has to be constructed reasonably. As shown in Section 2, exchangeability cannot yield the transition equation for the simple EA. This creates the research problem of finding a suitable modeling assumption to derive IPM. Therefore, in Section 4.2, we discuss the issue and propose to use i.i.d. as the modeling assumption in IPM.

4.1 Sufficient Conditions for Convergence of IPMs

To derive sufficient conditions for the convergence of the overall IPM, the core step is to derive conditions under which the operators in the IPM can be stacked and iterated.

As before, let EA_n and EA_∞ denote the EA with population size n and the IPM under analysis, respectively. Let Γ be an operator in the EA, and $\Gamma_n : \mathbb{M}^\infty \to \mathbb{M}^\infty$ and $\Gamma_\infty : \mathbb{M}^\infty \to \mathbb{M}^\infty$ be its corresponding expanded operators in EA_n and EA_∞, respectively. Note that Γ_n and Γ_∞ generate random elements of $\mathbb{S}^\infty$. To give an example, Γ_n and Γ_∞ may correspond to $\pi_n \circ \widetilde{G}_n$ and G_∞ in Figure 2, respectively.

We define a property under which Γ_∞ can be stacked with some other operator Ψ_∞ satisfying the same property without affecting the convergence of the overall IPM. In other words, for an EA using Ψ and Γ as its operators, we can prove the convergence of IPM by studying Ψ and Γ separately. We call this property "the stacking property." It is worth noting that if $\Phi = \Gamma$, then this property guarantees that Γ_∞ can be iterated for any number of times. Therefore, it also resolves the issue of iterating the algorithm.

Let $\mathbf{A}_\alpha$ be random elements in $\mathbb{M}^\infty$ for $\alpha \in \mathbb{N}_+ \cup \{\infty\}$. We have the following results.

DEFINITION 6 (The stacking property): *Given* $\mathbb{U} \subset \mathbb{M}^\infty$*, if for any converging sequence* $\mathbf{A}_n \xrightarrow{d} \mathbf{A}_\infty \in \mathbb{U}$*,* $\Gamma_n(\mathbf{A}_n) \xrightarrow{d} \Gamma_\infty(\mathbf{A}_\infty) \in \mathbb{U}$ *as* $n \to \infty$ *always holds, then we say that* Γ_∞ *has the stacking property on* $\mathbb{U}$*.*

THEOREM 6: *If* Ψ_∞ *and* Γ_∞ *have the stacking property on* $\mathbb{U}$*, then* $\Psi_\infty \circ \Gamma_\infty$ *has the stacking property on* $\mathbb{U}$*.*

PROOF: For any converging sequence $\mathbf{A}_n \xrightarrow{d} \mathbf{A}_\infty \in \mathbb{U} \subset \mathbb{M}^\infty$, because Γ_∞ has the stacking property on $\mathbb{U}$, we have $\Gamma_n(\mathbf{A}_n) \xrightarrow{d} \Gamma_\infty(\mathbf{A}_\infty) \in \mathbb{U}$. Then, $(\Gamma_n(\mathbf{A}_n))_n$ is also a converging sequence. Since Ψ_∞ has the stacking property on $\mathbb{U}$, then by definition we immediately have $(\Psi_n \circ \Gamma_n)(\mathbf{A}_n) \xrightarrow{d} (\Psi_\infty \circ \Gamma_\infty)(\mathbf{A}_\infty) \in \mathbb{U}$. □

By Theorem 6, any composition of Ψ_∞ and Γ_∞ has the stacking property on $\mathbb{U}$. In particular, $(\Gamma_\infty)^m$ has the stacking property on $\mathbb{U}$. The stacking property essentially guarantees that the convergence on $\mathbb{U}$ can be passed on to subsequent generations.

THEOREM 7 (Sufficient condition 1): *For an EA consisting of a single operator* Γ*, let* Γ *be modeled by* Γ_∞ *in the IPM,* EA_∞ *and* Γ_∞ *have the stacking property on some space* $\mathbb{U} \subset \mathbb{M}^\infty$*. If the initial populations of both EA and* EA_∞ *follow the same distribution* $P_\mathbf{X}$ *for some* $\mathbf{X} \in \mathbb{U}$*, then* EA_∞ *converges.*

PROOF: Note that for EA_n and EA_∞, the kth populations they generate are $(\Gamma_n)^k(\mathbf{X})$ and $(\Gamma_\infty)^k(\mathbf{X})$, respectively. By Theorem 6, $(\Gamma_\infty)^k$ has the stacking property on $\mathbb{U}$. Because the sequence $(\mathbf{X}, \mathbf{X}, \ldots)$ converges to $\mathbf{X} \in \mathbb{U}$, by Definition 6, $(\Gamma_n)^k(\mathbf{X}) \xrightarrow{d} (\Gamma_\infty)^k(\mathbf{X}) \in \mathbb{U}$ as $n \to \infty$. Since this holds for any $k \in \mathbb{N}$, by Definition 5, EA_∞ converges. □

By Theorems 6 and 7, we can prove the convergence of the overall IPM by proving that the operators in the IPM have the stacking property. Comparing with (30), it is clear that the stacking property is a sufficient condition. This is because the stacking property requires that $(\Gamma_n(\mathbf{A}_n))_n$ converges to a point in $\mathbb{U}$ for *any* converging sequence $(\mathbf{A}_n)_n$ satisfying $(\mathbf{A}_n)_n \xrightarrow{d} \mathbf{A}_\infty \in \mathbb{U}$, while (29) requires the convergence to hold only for the *specific* converging sequence $(\mathbf{Q}_k^n)_n$. Since $(\mathbf{Q}_k^n)_n$ is generated by the algorithm, it may have special characteristics regarding converging rate, distributions, etc. On the other hand, checking the stacking property may be easier than proving (29). This is because the stacking property is independent of the generation number k.

Another point worth discussing is the introduction of $\mathbb{U}$ in Definition 6. Of course, if we omit $\mathbb{U}$ (or equivalently let $\mathbb{U} = \mathbb{M}^\infty$), the stacking property will become "stronger" because if it holds, the convergence of the IPM is proved for the EA starting from *any* initial population. However, in that case the condition is so restricted that the stacking property cannot be proved for many operators.

In Definition 6, it is required that $\Gamma_n(\mathbf{A}_n) \overset{\mathrm{d}}{\to} \Gamma_\infty(\mathbf{A}_\infty) \in \mathbb{U}$ as $n \to \infty$. The sequence under investigation is $(\Gamma_n(\mathbf{A}_n))_n$, which is a sequence of changing operators $(\Gamma_n)_n$ on a sequence of changing inputs $(\mathbf{A}_n)_n$. As both the operators and the inputs change, the convergence of $(\Gamma_n(\mathbf{A}_n))_n$ may still be difficult to prove. Therefore, in the following, we further derive two sufficient conditions for the stacking property.

First, let $\mathbf{B}_{\alpha,\beta} = \Gamma_\beta(\mathbf{A}_\alpha)$, where $\alpha, \beta \in \mathbb{N}_+ \cup \{\infty\}$. Then, we have the following sufficient conditions for the stacking property.

THEOREM 8 (Sufficient condition 2): *For a space $\mathbb{U}$ and all converging sequences $\mathbf{A}_n \overset{\mathrm{d}}{\to} \mathbf{A}_\infty \in \mathbb{U}$, if the following two conditions*

1. $\exists M \in \mathbb{N}_+$, *such that for all* $m > M$, $\mathbf{B}_{n,m} \overset{\mathrm{d}}{\to} \mathbf{B}_{\infty,m}$ *uniformly as* $n \to \infty$, *i.e.* $\sup_{m>M} \rho_{\mathrm{d}}(\mathbf{B}_{n,m}, \mathbf{B}_{\infty,m}) \to 0$ *as* $n \to \infty$,

2. $\mathbf{B}_{\infty,m} \overset{\mathrm{d}}{\to} \mathbf{B}_{\infty,\infty} \in \mathbb{U}$ *as* $m \to \infty$,

are both met, then Γ_∞ has the stacking property on $\mathbb{U}$.

THEOREM 9 (Sufficient condition 3): *For a space $\mathbb{U}$ and all converging sequences $\mathbf{A}_n \overset{\mathrm{d}}{\to} \mathbf{A}_\infty \in \mathbb{U}$, if the following two conditions*

1. $\exists N \in \mathbb{N}_+$, *such that for all* $n > N$, $\mathbf{B}_{n,m} \overset{\mathrm{d}}{\to} \mathbf{B}_{n,\infty}$ *uniformly as* $m \to \infty$, *i.e.* $\sup_{n>N} \rho_{\mathrm{d}}(\mathbf{B}_{n,m}, \mathbf{B}_{n,\infty}) \to 0$ *as* $m \to \infty$,

2. $\mathbf{B}_{n,\infty} \overset{\mathrm{d}}{\to} \mathbf{B}_{\infty,\infty} \in \mathbb{U}$ *as* $n \to \infty$,

are both met, then Γ_∞ has the stacking property on $\mathbb{U}$.

Since Theorems 8 and 9 are symmetric in m and n, proving one of them leads to the other. In the following, we prove Theorem 8. Recall that ρ_{d} is the Prokhorov metric (Definition 4) and $\vee$ gets the maximal in the expression.

PROOF: $\forall \epsilon > 0$, by condition 1 in Theorem 8, $\exists N$ s.t. $\sup_{m>M} \rho_{\mathrm{d}}(\mathbf{B}_{n,m}, \mathbf{B}_{\infty,m}) < \frac{1}{2}\epsilon$ for all $n > N$. By condition 2 in Theorem 8, $\exists \widetilde{M}$ s.t. $\rho_{\mathrm{d}}(\mathbf{B}_{\infty,m}, \mathbf{B}_{\infty,\infty}) < \frac{1}{2}\epsilon$ for all $m > \widetilde{M}$. Now for all $l > M \vee N \vee \widetilde{M}$,

$$\rho_{\mathrm{d}}(\mathbf{B}_{l,l}, \mathbf{B}_{\infty,\infty}) \le \rho_{\mathrm{d}}(\mathbf{B}_{l,l}, \mathbf{B}_{\infty,l}) + \rho_{\mathrm{d}}(\mathbf{B}_{\infty,l}, \mathbf{B}_{\infty,\infty}) \le \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon.$$

Therefore, $\mathbf{B}_{n,n} \overset{\mathrm{d}}{\to} \mathbf{B}_{\infty,\infty}$ as $n \to \infty$. □

To understand these two theorems, consider the relationships between $\mathbf{A}_\alpha$ and $\mathbf{B}_{\alpha,\beta}$ illustrated by Figure 3. In the figure, the solid arrow represents the premise in Definition 6; that is, $\mathbf{A}_n \overset{\mathrm{d}}{\to} \mathbf{A}_\infty \in \mathbb{U}$ as $n \to \infty$. The double line arrow represents the direction to be proved for the stacking property on $\mathbb{U}$, i.e. $\mathbf{B}_{n,n} \overset{\mathrm{d}}{\to} \mathbf{B}_{\infty,\infty} \in \mathbb{U}$ as $n \to \infty$. The dashed arrows are the directions to be checked for Theorem 8 to hold. The wavy arrows are the directions to be checked for Theorem 9 to hold.
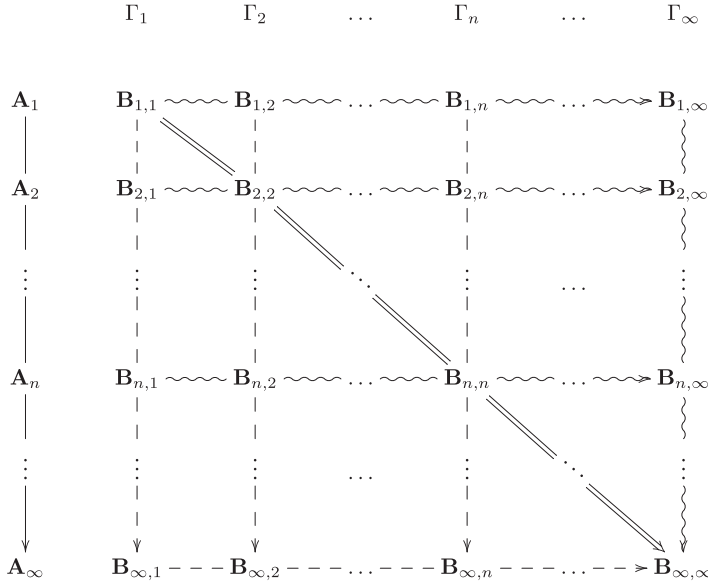
Figure 3: Relationships between $\mathbf{A}_\alpha$ and $\mathbf{B}_{\alpha,\beta}$.

Now it is clear that Theorems 8 and 9 bring benefits. For example, for Theorem 8, instead of proving the convergence for a sequence generated by changing operators and inputs ($\mathbf{B}_{n,n} \xrightarrow{\mathrm{d}} \mathbf{B}_{\infty,\infty}$), this sufficient condition considers the convergence of sequences generated by the *same* operator on changing inputs ($\mathbf{B}_{n,m} \xrightarrow{\mathrm{d}} \mathbf{B}_{\infty,m}$) and of the sequence generated by changing operators on the *same* input ($\mathbf{B}_{\infty,m} \xrightarrow{\mathrm{d}} \mathbf{B}_{\infty,\infty}$).

The reason we introduce M and N in Theorems 8 and 9, respectively, is to exclude some of the starting columns and rows in Figure 3, if necessary. This is useful in proving the convergence of the IPM of the k-ary recombination operator.

4.2 The I.I.D. Assumption

In this section, we address the issue of how to construct IPM. This issue also corresponds to how to choose the space $\mathbb{U}$ for the stacking property.

Before introducing the i.i.d. assumption, let us give an example. Consider the space $\mathbb{U} = \{\mathbf{x} \in \mathbb{M}^\infty | \mathrm{P}[\mathbf{x} = (c, c, \ldots)] = 1 \text{ for some } c \in \mathbb{S}\}$. If the initial population follows some distribution from $\mathbb{U}$, then the population consists of all identical individuals. If an EA with proportionate selection and crossover operates on this initial population, then all subsequent populations stay the same as the initial population. An IPM of this EA can be easily constructed, and it can be easily proved that the stacking property holds as long as the EA chooses its initial population from $\mathbb{U}$. However, this is not a very interesting case. This is because $\mathbb{U}$ is too small to model real EAs.

On the other hand, if $\mathbb{U} = \{\mathbf{x} \in \mathbb{M}^\infty | \mathbf{x} \text{ is exchangeable}\}$, $\mathbb{U}$ may be too big to derive meaningful results. This can be seen from our analysis in Section 2 which shows that under exchangeability it is not possible to derive transition equations of marginal distributions for the simple EA.

Therefore, choosing $\mathbb{U}$ should strike a balance between the capacity and the complexity of the IPM. In the following analysis, we choose $\mathbb{U}$ to be $\mathbb{U}_I = \{\mathbf{x} \in \mathbb{M}^\infty | \mathbf{x} \text{ is i.i.d.}\}$.

IPMs of EAs are constructed using the i.i.d. assumption, and we prove the convergence of the overall IPM by proving that the operators in the IPM have the stacking property on $\mathbb{U}_I$.

We choose $\mathbb{U}_I$ for the following reasons. First, in the real world, many EAs generate i.i.d. initial populations. Therefore this assumption is realistic. Secondly, i.i.d. random elements have the same marginal distributions. Therefore, IPM can be described by transition equations of marginal distributions. Finally, there is abundant literature on the converging laws and limit theorems of i.i.d. sequences. Therefore, the difficulty in constructing IPM can be greatly reduced compared with using other modeling assumptions.

In the following, we show how to construct IPM under the i.i.d. assumption. This process also relates to condition 2 in Theorem 8. It essentially describes how the IPM generates new populations.

Let the operator in the EA be Γ, and the corresponding operator in EA_m be Γ_m : $\mathbb{M}^\infty \to \mathbb{M}^\infty$. Recall that in our framework we only study EAs consisting of c.i.i.d. operators, therefore Γ_m generates c.i.i.d. outputs by using the first m elements of its input. The process that Γ_m generates *each* output can be described by the conditional p.d.f. $f_{\Gamma_m}(x|y_1, y_2, \ldots, y_m)$. Let $\mathbf{a} = (\mathbf{a}_i)_{i=1}^\infty \in \mathbb{M}^\infty$ be the input and $\mathbf{b} = (\mathbf{b}_i)_{i=1}^\infty = \Gamma_m(\mathbf{a})$ be the output, then the distribution of $\mathbf{b}$ can be completely described by its *finite-dimensional* p.d.f.s

$$f_{\pi_l(\mathbf{b})}(x_1, \ldots, x_l) = \int \int_{\mathbb{S}^m} \prod_{i=1}^l f_{\Gamma_m}(x_i|y_1, \ldots, y_m) \cdot f_{\pi_m(\mathbf{a})}(y_1, \ldots, y_m) \, dy_1 \ldots dy_m \qquad (30)$$

for every $l \in \mathbb{N}_+$.

To derive the IPM Γ_∞ for Γ, consider the case when $l = 1$ and $\mathbf{a} \in \mathbb{U}_I$ in (30). Noting that in this case $f_{\pi_m(\mathbf{a})}(y_1, \ldots, y_m) = \prod_{i=1}^m f_{\mathbf{a}_1}(y_i)$, we have

$$f_{\mathbf{b}_1}(x) = \int \int_{\mathbb{S}^m} f_{\Gamma_m}(x|y_1, \ldots, y_m) \cdot \prod_{i=1}^m f_{\mathbf{a}_1}(y_i) \, dy_1 \ldots dy_m. \qquad (31)$$

Now taking $m \to \infty$, (32) in the limit becomes the transition equation describing how Γ_∞ generates *each* new individual. Let the transition equation be

$$f_{\mathbf{b}_1} = T_\Gamma[f_{\mathbf{a}_1}], \qquad (32)$$

and let $\mathbf{c} = (\mathbf{c}_i)_{i=1}^\infty = \Gamma_\infty(\mathbf{a})$. Then how Γ_∞ generates l individuals can be described by the finite-dimensional p.d.f.s of $\mathbf{c}$:

$$f_{\pi_l(\mathbf{c})}(x_1, \ldots, x_l) = \prod_{i=1}^l T_\Gamma[f_{\mathbf{a}_1}](x_i) \qquad (33)$$

for every $l \in \mathbb{N}_+$. Overall, (34) describes the mapping $\Gamma_\infty : \mathbb{U}_I \to \mathbb{U}_I$.

To better understand the construction, it is important to realize that for Γ_∞ *both* the input and the output are i.i.d. In other words, Γ_∞ generates i.i.d. population dynamics to simulate the real population dynamics produced by Γ, only that the transition equation in Γ_∞ is derived by mimicking how Γ generates each new individual on i.i.d. inputs and taking the population size to infinity. In fact, if the stacking property on $\mathbb{U}_I$ is proved and the initial population is i.i.d., Γ_∞ will always take i.i.d. inputs and produce i.i.d. outputs. The behaviors of Γ_∞ on $\mathbb{U}_I$ are well-defined. On the other hand, $\Gamma_\infty(\mathbf{A} \notin \mathbb{U}_I)$

is not defined in the construction. This leaves us freedom. We can define $\Gamma_\infty(\mathbf{A} \notin \mathbb{U}_I)$ freely to facilitate proving the stacking property of Γ_∞. In particular, $\mathbf{B}_{n,\infty}$ for $n \in \mathbb{N}_+$ in Figure 3 can be defined freely to facilitate the analysis.

In fact, under the i.i.d. assumption, deriving the transition equation for most operators is the easy part. The more difficult part is to prove the stacking property of Γ_∞ on $\mathbb{U}_I$. To give an example, consider the transition equation (5) constructed in Qi and Palmieri (1994a), which models the joint effects of proportionate selection and mutation. As our analysis in Section 2 shows, it does not hold under the assumption of exchangeability. However, if the modeling assumption is i.i.d., the transition equation can be immediately proved (see our analysis in Section 2). This also applies to the transition equation built by the same authors for the uniform crossover operator (in Theorem 1 of Qi and Palmieri, 1994b), where the transition equation is in fact constructed under the i.i.d. assumption. Therefore, in the following analyses, we do not refer to the explicit form of the transition equation, unless it is needed. We only assume that the transition equation is successfully constructed, and it has the form (32) which is derived from (31) as $m \to \infty$.

The construction of the IPM also relates partly to condition 2 in Theorem 8. Comparing with this condition, it can be seen that for a successfully constructed Γ_∞, the following two facts are proved in the construction (m.p.w. stands for point-wise convergence of marginal p.d.f.s.).

1. $\mathbf{B}_{\infty,m} \xrightarrow{\text{m.p.w.}} \mathbf{B}_{\infty,\infty}$ as $m \to \infty$.

2. $\mathbf{B}_{\infty,\infty} \in \mathbb{U}_I$.

Of course, these two facts are not sufficient for this condition to hold. One still needs to prove $\mathbf{B}_{\infty,m} \xrightarrow{\text{d}} \mathbf{B}_{\infty,\infty}$ as $m \to \infty$. In other words, one has to consider convergence of finite dimensional distributions.

Finally, we sometimes use x for $x_1, \ldots, x_l$ if l is clear in the context. For example (30) can be rewritten as $f_{\pi_l(\mathbf{b})}(x) = \int \int_{\mathbb{S}^m} \prod_{i=1}^{l} f_{\Gamma_m}(x_i|y) \cdot f_{\pi_m(\mathbf{a})}(y) \, \mathrm{d}y$, where l takes m's place and means the population size which the operator is operating on.

5   Analysis of the Simple EA

In this section, we use the sufficient conditions to prove the convergence of IPMs for various simple EA operators. The operators of mutation and k-ary recombination are readily analyzed in Sections 5.1 and 5.2, respectively. In Section 5.3, we summarize this section and discuss our results.

5.1   Analysis of the Mutation Operator

Having derived sufficient conditions for the stacking property and constructed the IPM, we prove the convergence of the IPM of the mutation operator first. Mutation adds an i.i.d. random vector to each individual in the population. If the current population is $\mathbf{A} \in \mathbb{M}^\infty$, then the population after mutation satisfies $\mathcal{L}[\mathbf{B} = \Gamma_m(\mathbf{A})] = \mathcal{L}(\mathbf{A} + \mathbf{X})$ for all $m \in \mathbb{N}_+$, where $\mathbf{X} \in \mathbb{U}_I$ is a random element decided by the mutation operator. As the content of the mutation operator does not depend on m, we just write Γ to represent Γ_m. To give an example, $\mathbf{X}$ may be the sequence $(\mathbf{x}_1, \mathbf{x}_2, \ldots)$ with all $\mathbf{x}_i \in \mathbb{M}$ mutually independent and $\mathbf{x}_i \sim N(0, \mathrm{I}_d)$ for all $i \in \mathbb{N}_+$, where $N(a, B)$ is the multivariate normal distribution with mean a and covariance matrix B, and I_d is the d-dimensional identity matrix. Note that every time Γ is invoked, it generates perturbations independently.

For example, let $\mathbf{A}_1$ and $\mathbf{A}_2$ be two populations, then we can write $\Gamma(\mathbf{A}_i) = \mathbf{A}_i + \mathbf{X}_i$ for $i = 1, 2$ satisfying $\mathcal{L}(\mathbf{X}_1) = \mathcal{L}(\mathbf{X}_2) = \mathcal{L}(\mathbf{X})$ and $\{\mathbf{X}_i\}_{i=1,2}$ are mutually independent and independent from $\{\mathbf{A}_i\}_{i=1,2}$.

Next, consider Γ_∞. Recall that as an IPM, Γ_∞ simulates real population dynamics by taking i.i.d. inputs and producing i.i.d. outputs. If the marginal p.d.f.s of $\mathbf{A}$ and $\mathbf{X}$ are $f_\mathbf{a}$ and $f_\mathbf{x}$, respectively, then $\Gamma_\infty(\mathbf{A})$ generates i.i.d. individuals whose p.d.f.s are $f_\mathbf{a} * f_\mathbf{x}$, where $*$ stands for convolution. Given the construction, we can prove the stacking property of Γ_∞.

THEOREM 10 (Mutation): *Let Γ be the mutation operator, and Γ_∞ be the corresponding operator in the IPM constructed under the i.i.d. assumption, then Γ_∞ has the stacking property on $\mathbb{U}_\mathrm{I}$.*

PROOF: We use the notations and premises in Theorem 8. Refer to Figure 3. In particular, the sequence $(\mathbf{A}_n)$ and the limit $\mathbf{A}_\infty$ are given and $\mathbf{A}_n \overset{\mathrm{d}}{\to} \mathbf{A}_\infty \in \mathbb{U}_\mathrm{I}$ as $n \to \infty$.

Apparently,

$$[\Gamma_m(\mathbf{A}_\infty)]_{m=1}^\infty = [\Gamma(\mathbf{A}_\infty), \Gamma(\mathbf{A}_\infty), \ldots] \overset{\mathrm{d}}{\to} \Gamma(\mathbf{A}_\infty) = \Gamma_\infty(\mathbf{A}_\infty) \in \mathbb{U}_\mathrm{I}.$$

Therefore, condition 2 in Theorem 8 is satisfied.

Noting that condition 1 in Theorem 8 is equivalent to $\Gamma(\mathbf{A}_n) \overset{\mathrm{d}}{\to} \Gamma(\mathbf{A}_\infty)$, we prove this condition by proving that $\pi_i[\Gamma(\mathbf{A}_n)] \overset{\mathrm{d}}{\to} \pi_i[\Gamma(\mathbf{A}_\infty)]$ for all $i \in \mathbb{N}_+$. Then by Theorem 5, condition 1 in Theorem 8 holds. Then, as both conditions in Theorem 8 are satisfied, this theorem is proved.

Now, we prove $\pi_i[\Gamma(\mathbf{A}_n)] \overset{\mathrm{d}}{\to} \pi_i[\Gamma(\mathbf{A}_\infty)]$ for all $i \in \mathbb{N}_+$. First, note that $\Gamma(\mathbf{A}_\alpha) = \mathbf{A}_\alpha + \mathbf{X}_\alpha$ for all $\alpha \in \mathbb{N} \cup \{\infty\}$. $\{\mathbf{X}_\alpha \in \mathbb{M}^\infty\}$ are i.i.d. and independent from $\{\mathbf{A}_\alpha \in \mathbb{M}^\infty\}$. In addition, for every α, $\mathcal{L}(\mathbf{X}_\alpha) = \mathcal{L}(\mathbf{X})$.

Since $\mathcal{L}(\mathbf{X}_\alpha) = \mathcal{L}(\mathbf{X})$, it is apparent that $\mathbf{X}_n \overset{\mathrm{d}}{\to} \mathbf{X}_\infty$. Then by Theorem 5, we have $\pi_i(\mathbf{X}_n) \overset{\mathrm{d}}{\to} \pi_i(\mathbf{X}_\infty)$ and $\pi_i(\mathbf{A}_n) \overset{\mathrm{d}}{\to} \pi_i(\mathbf{A}_\infty)$.

Consider the product space $\mathbb{S}^i \times \mathbb{S}^i$. It is both separable and complete. Since $\pi_i(\mathbf{A}_\alpha)$ and $\pi_i(\mathbf{X}_\alpha)$ are independent, by Theorem 4, it follows that

$$\begin{bmatrix} \pi_i(\mathbf{A}_n) \\ \pi_i(\mathbf{X}_n) \end{bmatrix} \overset{\mathrm{d}}{\to} \begin{bmatrix} \pi_i(\mathbf{A}_\infty) \\ \pi_i(\mathbf{X}_\infty) \end{bmatrix}. \tag{34}$$

Note that

$$\pi_i[\Gamma(\mathbf{A}_\alpha)] = \begin{bmatrix} \mathrm{I} & \mathrm{I} \end{bmatrix} \begin{bmatrix} \pi_i(\mathbf{A}_\alpha) \\ \pi_i(\mathbf{X}_\alpha) \end{bmatrix} = h\left(\begin{bmatrix} \pi_i(\mathbf{A}_\alpha) \\ \pi_i(\mathbf{X}_\alpha) \end{bmatrix} \right), \tag{35}$$

where I is the identity matrix of appropriate dimension and $h : \mathbb{S}^i \times \mathbb{S}^i \to \mathbb{S}^i$ is a function satisfying $h(\begin{bmatrix} x \\ y \end{bmatrix}) = \begin{bmatrix} I & I \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$. Apparently h is continuous. Then by (34), (35), and Theorem 3, $\pi_i[\Gamma(\mathbf{A}_n)] \overset{\mathrm{d}}{\to} \pi_i[\Gamma(\mathbf{A}_\infty)]$ for any $i \in \mathbb{N}_+$. □

In the proof, we concatenate the input $(\mathbf{A}_n)$ and the randomness $(\mathbf{X}_n)$ of the mutation operator in a common product space, and represent Γ as a continuous function in that space. This technique is also used when analyzing other operators.

5.2 Analysis of *k*-ary Recombination

Consider the *k*-ary recombination operator and denote it by Γ. In EA_m, the operator is denoted by Γ_m. Γ_m works as follows. To generate a new individual, it first samples *k* individuals from the current *m*-sized population randomly *with* replacement. Assume

the current population consists of $\{\mathbf{x}_i\}_{x=1}^m$, and the selected k parents are $\{\mathbf{y}_i\}_{i=1}^k$, then $\{\mathbf{y}_i\}_{i=1}^k$ follows the probability:

$$\mathrm{P}(\mathbf{y}_i = \mathbf{x}_j) = \frac{1}{m} \text{ for all } i \in \{1, \dots, k\}, j \in \{1, \dots, m\}. \tag{36}$$

After the k parents are selected, Γ_m produces a new individual $\mathbf{x}$ following the formula

$$\mathbf{x} = \sum_{i=1}^k \mathbf{U}_i \mathbf{y}_i, \tag{37}$$

where $\{\mathbf{U}_i\}_{i=1}^k$ are random elements of $\mathbb{R}^{d \times d}$ (recall that $\mathbf{x}$ and $\mathbf{y}_i$ are random elements of $\mathbb{S} = \mathbb{R}^d$ modeling individuals in our framework). $\{\mathbf{U}_i\}_{i=1}^k$ are also independent of $\{\mathbf{y}_i\}_i$, and the *joint* distribution of $(\mathbf{U}_i)_i$ is decided by the inner mechanism of Γ. Overall, Γ_m generates the next population by repeatedly using this procedure to generate new individuals independently.

Our formulation seems strange at first sight, but it covers many real world recombination operators. For example, consider $k = 2$ and $\mathbf{U}_1 = \mathbf{U}_2 = \frac{1}{2} \mathbf{I}$. This operator is the crossover operator taking the mean of its two parents. On the other hand, if $k = 2$ and the distributions of $\mathbf{U}_1$ and $\mathbf{U}_2$ satisfy

$$\begin{cases} \mathbf{U}_1 = \mathrm{Diag}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_d) \\ \mathbf{U}_2 = \mathbf{I} - \mathbf{U}_1 \end{cases},$$

where Diag constructs a diagonal matrix from its inputs, $\{\mathbf{s}_i\}$ are i.i.d. random variables taking values in $\{0, 1\}$ satisfying $\mathrm{P}(\mathbf{s}_i = 0) = \mathrm{P}(\mathbf{s}_i = 1) = 1/2$, then this operator is the uniform crossover operator which sets value at each position from the two parents with probability $\frac{1}{2}$.

Consider the IPM Γ_∞. As stated in Section 4.2, we do not give the explicit form of the transition equation in Γ_∞. We assume that the IPM is successfully constructed, and the transition equation is derived by taking $m \to \infty$ in (31). The reason for this approach is not only because deriving the transition equation is generally easier than proving the convergence of the IPM, but also the formulation in (36) and (37) encompasses many real-world k-ary recombination operators. We do not delve into details of the mechanisms of these operators and derive a transition equation for each one of them. Instead, our approach is general in that as long as the IPM is successfully constructed, our analysis on the convergence of the IPM can always be applied.

The following theorem is the primary result of our analysis for the k-ary recombination operator.

THEOREM 11 (*k-ary recombination*): *Let Γ be the k-ary recombination operator, and Γ_∞ be the corresponding operator in the IPM constructed under the i.i.d. assumption, then Γ_∞ has the stacking property on $\mathbb{U}_\mathrm{I}$.*

PROOF: We use the notations and premises in Theorem 9. Refer to Figure 3. In particular, the sequence $(\mathbf{A}_n)$ and the limit $\mathbf{A}_\infty$ are given and $\mathbf{A}_n \xrightarrow{\mathrm{d}} \mathbf{A}_\infty \in \mathbb{U}_\mathrm{I}$ as $n \to \infty$.

We prove that

$$\pi_i(\mathbf{B}_{n,n}) \xrightarrow{\mathrm{d}} \pi_i(\mathbf{B}_{\infty,\infty}) \tag{38}$$

as $n \to \infty$ for any $i \in \mathbb{N}_+$. Then by Theorem 5, the conclusion follows.

The overall idea to prove (38) is that we first prove the convergence in distribution for the $k \cdot i$ selected parents; then because the recombination operator is continuous, (39) follows.

First, we decompose the operator $\pi_i \circ \Gamma_m : \mathbb{M}^\infty \to \mathbb{M}^i$. $\pi_i \circ \Gamma_m$ generates the i c.i.i.d. outputs one by one. This generation process can also be viewed as first selecting the i groups of k parents at once from the first m elements of the input (in total the intermediate output is $k \cdot i$ parents not necessarily distinct), then producing the i outputs one by one by using each group of k parents. In the following, we describe this process mathematically.

Consider $\Phi_m : \mathbb{M}^\infty \to \mathbb{M}^{k \cdot i}$. Let $\mathbf{x} = (\mathbf{x}_j)_{j=1}^\infty \in \mathbb{M}^\infty$ and $\mathbf{y} = (\mathbf{y}_j)_{j=1}^{k \cdot i} = \Phi_m(\mathbf{x})$. Let Φ_m be described by the probability

$$P(\mathbf{y}_j = \mathbf{x}_l) = \frac{1}{m} \text{ for all } j \in \{1, \ldots, k \cdot i\} \text{ and } l \in \{1, \ldots, m\}. \tag{39}$$

In essence, Φ_m describes how to select the $k \cdot i$ parents from $\mathbf{x}$.

Consider $\Psi : \mathbb{M}^{k \cdot i} \to \mathbb{M}^i$. Let

$$\mathbf{u} = (\mathbf{u}_{1,1}, \mathbf{u}_{1,2}, \ldots, \mathbf{u}_{1,k}, \mathbf{u}_{2,1}, \mathbf{u}_{2,2}, \ldots, \mathbf{u}_{2,k}, \ldots \ldots, \mathbf{u}_{i,1}, \mathbf{u}_{i,1}, \ldots, \mathbf{u}_{i,k}) \in \mathbb{M}^{k \cdot i}.$$

Let $\mathbf{v} = (\mathbf{v}_j)_{j=1}^i = \Psi(\mathbf{u})$. Let Ψ be described by

$$\mathbf{v}_j = \sum_{l=1}^k \mathbf{U}_{j,l} \mathbf{u}_{j,l} \text{ for all } j \in \{1, \ldots, i\} \tag{40}$$

in which $\mathcal{L}[(\mathbf{U}_{j,l})_{l=1}^k] = \mathcal{L}[(\mathbf{U}_l)_{l=1}^k]$, where $\{\mathbf{U}_l\}$ are decided by the recombination operator Γ as in (37), and $(\mathbf{U}_{j,l})_{l=1}^k$ are independent for different j. In essence, Ψ describes how to generate the i individuals from the $k \cdot i$ parents.

Now it is obvious that $\pi_i \circ \Gamma_m = \Psi \circ \Phi_m$. Therefore,

$$\pi_i(\mathbf{B}_{m,\alpha}) = (\pi_i \circ \Gamma_m)(\mathbf{A}_\alpha) = (\Psi \circ \Phi_m)(\mathbf{A}_\alpha) \tag{41}$$

for all $m \in \mathbb{N}_+$ and $\alpha \in \mathbb{N}_+ \cup \{\infty\}$.

Next, consider $\pi_i \circ \Gamma_\infty : \mathbb{M}^\infty \to \mathbb{M}^i$. Let $\Phi_\infty = \pi_{k \cdot i}$, we prove that

$$\mathcal{L}[(\pi_i \circ \Gamma_\infty)(\mathbf{A})] = \mathcal{L}[(\Psi \circ \Phi_\infty)(\mathbf{A})], \forall \mathbf{A} \in \mathbb{U}_I. \tag{42}$$

Eq. (42) is almost obvious because both operators generate i.i.d. outputs, and both marginal p.d.f.s of the outputs follow the same distribution decided by Ψ on k i.i.d. parents from $\mathbf{A}$. In other words, $\Psi \circ \Phi_\infty$ is a model of $\pi_i \circ \Gamma_\infty$ on i.i.d. inputs. The outputs they generate on the same i.i.d. input follow the same distribution.

Since $\mathbf{A}_\infty \in \mathbb{U}_I$, by (42),

$$\mathcal{L}[\pi_i(\mathbf{B}_{\infty,\infty}) = (\pi_i \circ \Gamma_\infty)(\mathbf{A}_\infty)] = \mathcal{L}[(\Psi \circ \Phi_\infty)(\mathbf{A}_\infty)]. \tag{43}$$

Then (38) is equivalent to

$$(\Psi \circ \Phi_n)(\mathbf{A}_n) \xrightarrow{d} (\Psi \circ \Phi_\infty)(\mathbf{A}_\infty). \tag{44}$$

as $n \to \infty$ for any $i \in \mathbb{N}_+$.

To prove (44), we prove the following two conditions.

1. $\exists N \in \mathbb{N}_+$, such that for all $n > N$, $\Phi_m(\mathbf{A}_n) \xrightarrow{d} \Phi_\infty(\mathbf{A}_n)$ uniformly as $m \to \infty$; that is, $\sup_{n>N} \rho_d[\Phi_m(\mathbf{A}_n), \Phi_\infty(\mathbf{A}_n)] \to 0$ as $m \to \infty$.

2. $\Phi_\infty(\mathbf{A}_n) \xrightarrow{d} \Phi_\infty(\mathbf{A}_\infty)$ as $n \to \infty$ and $\Phi_\infty(\mathbf{A}_\infty)$ is i.i.d.

These two conditions correspond to the conditions in Theorem 9. Since Φ_α is from $\mathbb{M}^\infty$ to $\mathbb{M}^{k \cdot i}$, we cannot directly apply Theorem 9. However, it is easy to extend the proof of Theorem 9 to prove that these two conditions lead to $\Phi_n(\mathbf{A}_n) \xrightarrow{d} \Phi_\infty(\mathbf{A}_\infty)$ as $n \to \infty$. Then, by (40) it is apparent that Ψ is a continuous function of its input and inner randomness. By concatenating the input and the inner randomness using the same technique as that used in the proof for Theorem 10, (44) can be proved. Then this theorem is proved.

In the remainder of the proof, we prove conditions 1 and 2. These conditions can be understood by replacing the top line with Φ_m in Figure 3. $\qquad\square$

Proof of Condition 2

Since $\Phi_\infty = \pi_{k \cdot i} : \mathbb{S}^\infty \to \mathbb{S}^{k \cdot i}$ (recall that $\pi_{k \cdot i}$ can be viewed both as a mapping from $\mathbb{S}^\infty$ to $\mathbb{S}^{k \cdot i}$ and from $\mathbb{M}^\infty$ to $\mathbb{M}^{k \cdot i}$), Φ_∞ is continuous (see Example 1.2 in Billingsley, 1999). Since $\mathbf{A}_n \xrightarrow{d} \mathbf{A}_\infty$, by Theorem 3, $\Phi_\infty(\mathbf{A}_n) \xrightarrow{d} \Phi_\infty(\mathbf{A}_\infty)$. Apparently, $\Phi_\infty(\mathbf{A}_\infty)$ is i.i.d. Therefore condition 2 is proved.

It is worth noting that this simple proof comes partly from our extension of $\Psi \circ \Phi_\infty$ to inputs $\mathbf{A} \notin \mathbb{U}_I$. In fact, the only requirement for Φ_∞ is (42); that is, $\Psi \circ \Phi_\infty$ should model $\pi_i \circ \Gamma_\infty$ on *i.i.d.* inputs. By defining Φ_∞ to be $\pi_{k \cdot i}$, it can take non-i.i.d. inputs such as $\mathbf{A}_n$. Thus this condition can be proved. In Figure 3, this corresponds to our freedom of defining $\mathbf{B}_{n,\infty}, n \in \mathbb{N}_+$.

Proof of Condition 1

To prove condition 1, we first give another representation of $\Phi_m(\mathbf{A}_\alpha)$, where $m > k \cdot i$ and $\alpha \in \mathbb{N}_+ \cup \{\infty\}$. This representation is based on the following mutually exclusive cases.

1. The $k \cdot i$ parents chosen from $\mathbf{A}_\alpha$ by Φ_m are distinct.

2. There are duplicates in the $k \cdot i$ parents which are chosen from $\mathbf{A}_\alpha$ by Φ_m.

Let $\mathbf{s}_{m,\alpha}$ be random variables taking values in $\{0, 1\}$, with probability

$$
\begin{aligned}
p(m) &= \mathrm{P}(\mathbf{s}_{m,\alpha} = 1) \\
&= \mathrm{P}(\Phi_m \text{ chooses } k \cdot i \text{ distinct parents from } \mathbf{A}_\alpha) \\
&= \frac{m \cdot (m-1) \cdot \cdots \cdot (m - k \cdot i + 1)}{m^{k \cdot i}}.
\end{aligned}
\tag{45}
$$

Let $\mathbf{x}_{m,\alpha} \in \mathbb{M}^{k \cdot i}$ follow the *conditional* distribution of the $k \cdot i$ parents when $\mathbf{s}_{m,\alpha} = 1$, and $\mathbf{y}_{m,\alpha} \in \mathbb{M}^{k \cdot i}$ follow the *conditional* distribution of the $k \cdot i$ parents when $\mathbf{s}_{m,\alpha} = 0$, then $\Phi_m(\mathbf{A}_\alpha)$ can be further represented as

$$
\Phi_m(\mathbf{A}_\alpha) = \mathbf{s}_{m,\alpha} \cdot \mathbf{x}_{m,\alpha} + (1 - \mathbf{s}_{m,\alpha}) \cdot \mathbf{y}_{m,\alpha}.
\tag{46}
$$

For our purpose, it is not necessary to explicitly describe the distribution of $\mathbf{x}_{m,\alpha}$ and $\mathbf{y}_{m,\alpha}$. The only useful fact is that by exchangeability of $\mathbf{A}_\alpha$,

$$
\mathcal{L}(\mathbf{x}_{m,\alpha}) = \mathcal{L}[\Phi_\infty(\mathbf{A}_\alpha)].
\tag{47}
$$

To put it another way, $\mathbf{x}_{m,\alpha}$ and $\Phi_\infty(\mathbf{A}_\alpha)$ both follow the same distribution of $k \cdot i$ *distinct* individuals from the current *exchangeable* population $\mathbf{A}_\alpha$. Also note that $\{\mathbf{s}_{m,\alpha}\}_\alpha$ are i.i.d. random variables. They are independent of $\mathbf{x}_{m,\alpha}$ and $\mathbf{y}_{m,\alpha}$.

Now consider $\mathrm{P}[\Phi_m(\mathbf{A}_n) \in A]$ for any $A \in \mathcal{S}^{k \cdot i}$. By conditioning on whether the $k \cdot i$ parents are distinct, we have

$$\mathrm{P}[\Phi_m(\mathbf{A}_n) \in A] = p(m) \cdot \mathrm{P}(\mathbf{x}_{m,n} \in A) + [1 - p(m)] \cdot \mathrm{P}(\mathbf{y}_{m,n} \in A).$$

Then by (47),

$$\begin{aligned}
&\mathrm{P}[\Phi_m(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] \\
&= [p(m) - 1] \cdot \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] + [1 - p(m)] \cdot \mathrm{P}(\mathbf{y}_{m,n} \in A). \quad (48)
\end{aligned}$$

Since $p(m)$, $\mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A]$ and $\mathrm{P}(\mathbf{y}_{m,n} \in A)$ are all less than or equal to 1,

$$\begin{aligned}
p(m) - 1 &\le [p(m) - 1] \cdot \mathrm{P}(\Phi_\infty(\mathbf{A}_n) \in A) \\
&\le \mathrm{P}[\Phi_m(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] \\
&\le [p(m) - 1] \cdot \mathrm{P}(\Phi_\infty(\mathbf{A}_n) \in A) + [1 - p(m)] \\
&\le 1 - p(m),
\end{aligned}$$

i.e. $\left| \mathrm{P}[\Phi_m(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] \right| \le 1 - p(m)$ for all A. Taking supremum over all A, we have

$$\sup_{A \in \mathcal{S}^{k \cdot i}} \left| \mathrm{P}[\Phi_m(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] \right| \le 1 - p(m) \quad (49)$$

The left-hand side of (49) is the total variation distance between $\Phi_m(\mathbf{A}_n)$ and $\Phi_\infty(\mathbf{A}_n)$. It is an upper bound of the Prokhorov distance (see Gibbs and Su, 2002 for its definition and properties). Since the bound $1 - p(m)$ is uniform with respect to n and $p(m) \to 1$ as $m \to \infty$, we have

$$\sup_n \rho_d[\Phi_m(\mathbf{A}_n), \Phi_\infty(\mathbf{A}_n)] \le 1 - p(m) \to 0 \text{ as } m \to \infty. \quad (50)$$

This is exactly condition 1. Therefore this theorem is proved.

Or, if we do not want to use the total variance distance, we have the following result for any $\Phi_\infty(\mathbf{A}_\infty)$-continuity set $A \in \mathbb{S}^{k \cdot i}$.

$$\begin{aligned}
&\left| \mathrm{P}[\Phi_n(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_\infty) \in A] \right| \\
&\le \left| \mathrm{P}[\Phi_n(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] \right| + \left| \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_\infty) \in A] \right| \\
&\le 1 - p(n) + \left| \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_\infty) \in A] \right|. \quad (51)
\end{aligned}$$

Since we already proved $\Phi_\infty(\mathbf{A}_n) \xrightarrow{\text{d}} \Phi_\infty(\mathbf{A}_\infty)$, by 4) in Theorem 2, $\left| \mathrm{P}[\Phi_\infty(\mathbf{A}_n) \in A] - \mathrm{P}[\Phi_\infty(\mathbf{A}_\infty) \in A] \right| \to 0$. Then apparently (51) converges to 0. Noting that A is arbitrary, by applying 4) in Theorem 2 again, $\Phi_n(\mathbf{A}_n) \xrightarrow{\text{d}} \Phi_\infty(\mathbf{A}_\infty)$ is proved. □

We give a brief discussion of the proof. In our opinion, the most critical step of our proof is decomposing the k-ary recombination operator to two suboperators: one is responsible for selecting parents (Φ), the other is responsible for combining them (Ψ). In addition, for parent selection, the suboperator does *not* use the information of fitness values. Rather, it selects parents "blindly" according to its own rules (uniform sampling with replacement). This makes the operator Φ easier to analyze because the way it selects parents does not rely on its input. Therefore, we can prove uniform convergence in (50).

Another point worth mentioning is the choice of Theorem 9 in our proof. Though Theorems 8 and 9 are symmetric, the difficulties of proving them are quite different. In fact, it is very difficult to prove the uniform convergence condition in Theorem 8.

Finally, our proof can be easily extended to cover *k*-ary recombination operators using uniform sampling *without* replacement to select parents for each offspring. The overall proof framework roughly stays the same.

5.3   Summary

In this section, we analyzed the simple EA within the proposed framework. As the analysis shows, although the convergence of IPM is rigorously defined, actually proving the convergence for operators usually takes a lot of effort. We did analysis under the IPM construction and sufficient conditions from Section 4, and used various techniques to analyze the mutation operator and the *k*-ary recombination operator. It can be seen that although the sufficient conditions can provide general directions for the proofs, there are still many details to be worked out in order to analyze different operators.

To appreciate the significance of our work, it is worth noting that in Qi and Palmieri (1994a,b), the convergence of the IPMs of the mutation operator, the uniform crossover operator and the proportionate selection operator was not properly proved, and the issue of stacking of operators and iterating the algorithm was not addressed at all. In this article, however, we have proved the convergence of IPMs of several general operators. Since these general operators cover the operators studied in Qi and Palmieri (1994a,b) as special cases, the convergence of the IPMs of mutation and uniform crossover are actually proved in this article. Besides, our proof does not depend on the explicit form of the transition equation of the IPM. As long as the IPM is constructed under the i.i.d. assumption, our proof is valid.

As a consequence of our result, consider the explicit form of the transition equation for the uniform crossover operator derived in Section II in Qi and Palmieri (1994b). As the authors' proof was problematic and incomplete, the derivation of the transition equation was not well founded. However, it can be seen that the authors' derivation is in fact equivalent to constructing the IPM under the i.i.d. assumption. Since we have already proved the convergence of IPM of the *k*-ary crossover operator, the analysis in Qi and Palmieri (1994b) regarding the explicit form of the transition equation can be retained.

6   Conclusion and Future Research

In this article, we revisited the existing literature on the theoretical foundations of IPMs, and proposed an analytical framework for IPMs based on convergence in distribution for random elements taking values in the metric space of infinite sequences. Under the framework, commonly used operators such as mutation and recombination were analyzed. Our approach and analyses are new. There are many topics worth studying for future research.

Perhaps the most immediate topic is to analyze the proportionate selection operator in our framework. The reason that the mutation operator and the *k*-ary recombination operator can be readily analyzed is partly because they do not use the information of the fitness value. Also to generate a new individual, these operators draw information from a fixed number of parents. On the other hand, to generate each new individual, the proportionate selection operator actually gathers and uses fitness values of the whole population. This makes analyzing proportionate selection difficult.

We think further analysis on proportionate selection can be conducted in the following two directions.

1. In the analyses we tried to prove the stacking property on $\mathbb{U}_I$ for the IPM of proportionate selection. Apart from more efforts trying to prove/disprove this property, it is worth considering modifying the space $\mathbb{U}_I$. For example, we can incorporate the *rate* of convergence into the space. If we can prove the stacking property on $\mathbb{U}_I \cap \mathbb{U}$ where $\mathbb{U}$ is the space of converging sequences with rate $O(h(n))$, it is also a meaningful result.

2. Another strategy is to bypass the sufficient conditions and return to Definition 5 to prove $\mathbf{Q}_k^n \overset{d}{\to} \mathbf{Q}_k^\infty$ for every k. This is the original method. In essence, it requires studying the convergence of nesting integrals.

Apart from proportionate selection, it is also worth studying whether other operators, such as ranking selection, can be analyzed in our framework. As many of these operators do not generate c.i.i.d. offspring, it makes deriving the IPM and proving its convergence difficult, if not impossible. In this regard, we believe new techniques of modeling and extensions of the framework are fruitful directions for further research.

Finally, it is possible to extend the concept of "incidence vectors" proposed by Vose to the continuous search space. After all, as noted by Vose himself, incidence vectors can also be viewed as marginal p.d.f.s of individuals. As a consequence, the cases of EAs on discrete and continuous solution spaces indeed do bear some resemblance. By an easy extension, the incidence vectors in the continuous space can be defined as functions with the form $\sum c_i \delta(x_i)$, where δ is the Dirac function and c_i is the rational number representing the fraction that x_i appears in the population. If similar analyses based on this extension can be carried out, many results in Nix and Vose (1992) and Vose (1999b,a, 2004) can be extended to the continuous space.

References

Billingsley, P. (1999). *Convergence of probability measures*. 2nd ed. New York: John Wiley & Sons.

Dudley, R. M. (2002). *Real analysis and probability*. 2nd ed. Cambridge: Cambridge University Press.

Gibbs, A. L., and Su, F. E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, 70(3): 419–435.

Kallenberg, O. (2002). *Foundations of modern probability*. 2nd ed. New York: Springer.

Nix, A., and Vose, M. D. (1992). Modeling genetic algorithms with Markov chains. *Annals of Mathematics and Artificial Intelligence*, 5(1): 79–88.

Port, S. C. (1994). *Theoretical probability for applications*. New York: John Wiley & Sons.

Qi, X., and Palmieri, F. (1994a). Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space. Part I: Basic properties of selection and mutation. *IEEE Transactions on Neural Networks*, 5(1): 102–119.

Qi, X., and Palmieri, F. (1994b). Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space. Part II: Analysis of the diversification role of crossover. *IEEE Transactions on Neural Networks*, 5(1): 120–129.

Taylor, R. L., Daffer, P. Z., and Patterson, R. F. (1985). *Limit theorems for sums of exchangeable random variables*. Totowa, NJ: Rowman & Allanheld.

Vose, M. D. (1999a). *The simple genetic algorithm: Foundations and theory*. Cambridge, MA: MIT Press.

Vose, M. D. (1999b). What are genetic algorithms? A mathematical prespective. In L. Davis, K. De Jong, M. Vose, and L. Whitley (Eds.), *Evolutionary algorithms*, pp. 251–276. The IMA Volumes in Mathematics and Its Applications, vol. 111. New York: Springer.

Vose, M. D. (2004). Infinite population GA tutorial. Technical Report ut-cs-04-533. The University of Tennessee, Knoxville.

Yong, G., Xiaofeng, Q., and Palmieri, F. (1998). Comments on "Theoretical analysis of evolutionary algorithms with an infinite population size in continuous space. I. Basic properties of selection and mutation" [with reply]. *IEEE Transactions on Neural Networks*, 9(2): 341–343.