

Continuous Lunches are Free!

Anne Auger
TAO Team - INRIA Futurs
LRI - Paris-Sud University
91405 Orsay Cedex, France
Anne.Auger@inria.fr

Olivier Teytaud
TAO Team - INRIA Futurs
LRI - Paris-Sud University
91405 Orsay Cedex, France
Olivier.Teytaud@inria.fr

ABSTRACT

This paper investigates extensions of No Free Lunch (NFL) theorems to countably infinite and uncountable infinite domains. The original NFL due to Wolpert and Macready states that all search heuristics have the same performance when averaged over the uniform distribution over all possible functions. For infinite domains the extension of the concept of distribution over all possible functions involves measurability issues and stochastic process theory. For countably infinite domains, we prove that the natural extension of NFL theorems does not hold, but that a weaker form of NFL does hold, by stating the existence of non-trivial distributions of fitness leading to equal performance for all search heuristics. Our main result is that for continuous domains, NFL does not hold.

Categories and Subject Descriptors

G.1.6 [Numerical Analysis]: Optimization—*Global optimization, Unconstrained optimization*; F.2.1 [Analysis of Algorithms and Problem Complexity]: [Numerical Algorithms and Problems]

General Terms

Algorithms, Theory

Keywords

Free-Lunch, No-Free-Lunch, Kolmogorov's extension Theorem

1. INTRODUCTION

Search heuristics like evolutionary algorithms, tabu search, grid search are general heuristics that can be applied to any objective function. Much research is devoted to develop heuristics that are superior to others when the target functions belong to a certain class of problems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GECCO'07, July 7–11, 2007, London, England, United Kingdom.
Copyright 2007 ACM 978-1-59593-697-4/07/0007 ...\$5.00.

The No-Free-Lunch (NFL) theorem for optimization [13], introduced for fitness functions defined on a finite domain, rules out statements that some search heuristics have some advantages when performances are averaged uniformly among all possible functions. Controversial discussions on the usefulness of search heuristics ensued from this Theorem [3, 10]. Droste et al. argue that the NFL scenario is not a realistic one and they show that for realistic black box scenarios, NFL will not hold [7]. Igel and Toussaint [8] show that classes of functions relevant in practice are not likely to satisfy the (sharpened) NFL scenario [11]. Corne and Knowles [2] show that NFL does not hold in the multi-objective case.

Regardless of those results, the NFL theorem is still frequently cited to weaken progress made by practitioners or theoreticians. Besides it is quite common to hear people using the “ultimate” NFL argument for search heuristics on a continuous search space.

However, a basic assumption of NFL theorems is that the search space is finite. In that case “averaging uniformly among all possible functions” has a clear meaning: the average is made with respect to the cardinal and uniform random fitness can be easily defined as a mapping from a probability space to the finite set of all possible functions. It is more tricky to give a meaning to uniform average among functions when the domain is countably infinite or continuous (uncountable infinite). More generally, it is difficult to define a random mapping from a probability space to the (infinite) space of all possible functions. Doing so in a proper way involves measurability and stochastic process theory [4, 9, 1].

In this paper, we investigate NFL results for countably infinite and continuous domains. In Section 2 we start by reviewing NFL theorems for finite domains. We recall important definitions from measure theory required when dealing with uncountable infinite domains. We also introduce definitions required for the extensions of (N)FL theorems to countably infinite and uncountable domains. In Section 3 we show that the natural extension of NFL results for countably infinite domain does not hold but show that a weaker form of NFL does hold. In Section 4 we show the main result of this paper, namely that NFL does not hold for continuous domains.

2. PRELIMINARIES

In this section, we present (i) the finite case (ii) the definitions required for the sequel.

2.1 Finite lunches

We present in this section NFL theorems for objective functions mapping a finite domain \mathcal{X} , with cardinal $|\mathcal{X}|$, into a finite codomain $\mathcal{Y} \subset \mathbb{R}$ [13, 11, 8, 7, 2]. The search heuristics considered are randomized or deterministic and it is assumed that they are non-repeating. In practice, this can be ensured by archiving the different inputs. For any integer m in $\{1, \dots, |\mathcal{X}|\}$, the vector (x_1, \dots, x_m) represents the m first iterates of a search algorithm and the vector $(f(x_1), \dots, f(x_m))$ their associated objective values for a given function f mapping \mathcal{X} into \mathcal{Y} . The performance of an algorithm a after m iterations is measured using the vector

$$Y(f, m, a) = \langle f(x_1), \dots, f(x_m) \rangle \quad (1)$$

where we stick to the notations used in [8]. Let c denotes a performance measure mapping the vector of cost-values to the real numbers. The function c can be for instance the minimum value of the vector $Y(f, m, a)$ or the number of iterations before reaching a given value.

We denote $\Pi(\mathcal{X})$ the set of permutations on \mathcal{X} . A set of functions \mathcal{F} is closed under permutation (c.u.p.) if for any $f \in \mathcal{F}$ and any permutation $\pi \in \Pi(\mathcal{X})$, $f \circ \pi \in \mathcal{F}$.

The original NFL for optimization was stated for the set of all possible functions on \mathcal{X} , *i.e.* $\mathcal{F} = \mathcal{Y}^{\mathcal{X}}$ [13] and has been generalized for c.u.p. subsets [11, 7]:

THEOREM 2.1 (NFL FOR C.U.P. SUBSETS). *Let \mathcal{F} be a subset of $\mathcal{Y}^{\mathcal{X}}$. Then, for any two algorithms a and b , any m in $\{1, \dots, |\mathcal{X}|\}$, any performance measure c , any $k \in \mathbb{R}$*

$$\sum_{f \in \mathcal{F}} \delta(k, c(Y(f, m, a))) = \sum_{f \in \mathcal{F}} \delta(k, c(Y(f, m, b))) \quad (2)$$

*iff \mathcal{F} is c.u.p.*¹

In the previous theorem, averaging is done by uniform summation over \mathcal{F} which implicitly means that all functions in \mathcal{F} are equally likely. An equivalent point of view is to define f as a random variable taking values in the set of function \mathcal{F} , with the same probability for each function in \mathcal{F} . The vector $Y(f, m, a)$ defined in Eq. 1 is then a random vector and an equivalent formulation to “for all k in \mathbb{R} , Eq. 2 holds” is

$$\forall k \in \mathbb{R}, \quad \mathbb{P}(c(Y(f, m, a)) = k) = \mathbb{P}(c(Y(f, m, b)) = k) \quad (3)$$

where the random variable f is uniformly distributed on \mathcal{F} , *i.e.* for any $f_0 \in \mathcal{F}$, $\mathbb{P}(f = f_0)$ is constant equal to $1/|\mathcal{F}|$. Eq. 3 is an equivalent way to state that the distribution of $c(Y(f, m, a))$ and $c(Y(f, m, b))$ are the same and an equivalent formulation for Theorem 2.1 is

THEOREM 2.2 (NFL FOR C.U.P. SUBSETS). *Let \mathcal{F} be a subset of $\mathcal{Y}^{\mathcal{X}}$ and f a random variable uniformly distributed on \mathcal{F} . Then, for any two algorithms a and b , any m in $\{1, \dots, |\mathcal{X}|\}$ and any performance measure c , $c(Y(f, m, a))$ and $c(Y(f, m, b))$ follow the same distribution iff \mathcal{F} is c.u.p.*

A generalization of this NFL theorem to non uniform distributions of fitness has been presented in [8]. They consider the histogram h_{f_0} of a function $f_0 \in \mathcal{Y}^{\mathcal{X}}$ defined for each $y \in \mathcal{Y}$ as the cardinal of the preimage $f_0^{-1}(y)$, *i.e.*

$$h_{f_0} : y \in \mathcal{Y} \rightarrow |f_0^{-1}(y)|$$

¹The Kronecker delta function, δ is defined as $\delta(a, b) = 1$ if $a = b$ and $\delta(a, b) = 0$ otherwise.

Then, Theorem 2.2 holds for random variables f having distributions constant for functions having the same histogram [8], *i.e.*

$$\text{if } h_{f_0} = h_{f_1}, \mathbb{P}(f = f_0) = \mathbb{P}(f = f_1). \quad (4)$$

One simple random variable satisfying the condition in Eq. 4 is defined as follows: consider the c.u.p. subset $\mathcal{F}_{f_0}^{\Pi}$ defined for a function $f_0 \in \mathcal{Y}^{\mathcal{X}}$ as

$$\mathcal{F}_{f_0}^{\Pi} = \{f_0 \circ \pi \text{ for } \pi \in \Pi(\mathcal{X})\} \quad (5)$$

and define the corresponding random variable $f_{\mathcal{F}_{f_0}^{\Pi}}$ as

$$f_{\mathcal{F}_{f_0}^{\Pi}} = f_0 \circ \pi \text{ where } \pi \text{ is a random variable uniformly distributed on } \Pi(\mathcal{X}). \quad (6)$$

Then $f_{\mathcal{F}_{f_0}^{\Pi}}$ satisfies the histogram condition (Eq. 4) and also Theorem 2.2. Averaging with any unequal weights multiple random variables of this form (with different functions f_0) provides a random variable satisfying Eq. 4 without, in general, satisfying the uniformity on a c.u.p. subset condition required in Theorem 2.2.

2.2 Generalization

The generalization of NFL theorems for countably infinite and uncountable infinite domains calls upon measurability theory. We summarize here some basics that we will need in the sequel. Consider a probability space $(\Omega, \mathfrak{A}, \mathbb{P})$, where \mathfrak{A} is a σ -algebra on Ω and \mathbb{P} a probability measure defined on \mathfrak{A} . When \mathcal{X} is finite, a mapping $X : \Omega \rightarrow \mathcal{X}$ is a random variable² if for any x_0 , $\{\omega \in \Omega, X(\omega) = x_0\}$, denoted in short $\{X = x_0\}$, is in \mathfrak{A} . $\mathbb{P}_X : x_0 \mapsto \mathbb{P}_X(x_0) = \mathbb{P}(\{X = x_0\})$ is then well defined and called the distribution function of X . This definition is extended to $\mathcal{X} = \mathbb{R}^n$ using measurability. A mapping $X : \Omega \rightarrow \mathbb{R}^n$ is a random variable if it is a measurable function, *i.e.* a function such that $X^{-1}(E) = \{\omega \in \Omega, X(\omega) \in E\}$ is in \mathfrak{A} for any measurable subset E of \mathbb{R}^n . This implies that $\mathbb{P}_X(E) = \mathbb{P}(\{\omega \in \Omega; X(\omega) \in E\})$ is defined for any measurable $E \subset \mathbb{R}^n$. When \mathcal{X} has no natural measure, defining something similar to a random variable is more difficult. For defining a random variable with values in $\mathcal{Y}^{\mathcal{X}}$ with $\mathcal{Y} \subset \mathbb{R}$ and \mathcal{X} finite, one can simply use the natural isomorphism from $\mathcal{Y}^{\mathcal{X}}$ to \mathbb{R}^n where n is the cardinal of \mathcal{X} . This induces a measure on $\mathcal{Y}^{\mathcal{X}}$, and this is the usual measure in this case. When \mathcal{X} is infinite but countable, then Kolmogorov’s extension theorem provides a natural extension. But when \mathcal{X} is uncountable, typically $\mathcal{X} = \mathbb{R}^n$, we need stochastic processes or random fields [12]³. This will be detailed in Section 4.

Performance measurement

In Theorem 2.2, performance is measured using the distribution of $c(Y(f, m, a))$: two algorithms a and b are equivalent if the distributions of $c(Y(f, m, a))$ and $c(Y(f, m, b))$ are the same. Lemma 2.3 below shows that it is equivalent to requiring that the distributions of $Y(f, m, a)$ and $Y(f, m, b)$ are the same.

²We implicitly use here the standard σ -algebra on finite sets. With other σ -algebra, more restrictive definitions of a random variable could be stated.

³Stochastic processes and random fields are formally very similar; however, the term “random field” is usually preferred when the dimension of \mathcal{X} is greater than 1.

LEMMA 2.3. Let \mathcal{M} be the set of measurable functions from \mathbb{R}^n to \mathbb{R} and \mathcal{M}' the set of characteristic functions⁴ of measurable sets of the form $]-\infty, t_1] \times]-\infty, t_2] \times \cdots \times]-\infty, t_n] \subset \mathbb{R}^n$. Then, for any family \mathcal{A} of random variables in \mathbb{R}^n , the following equations are equivalent:

$$\begin{aligned} \forall (a_1, \dots, a_n) \in \mathcal{A}, \forall c \in \mathcal{M}, \\ c(a_1, \dots, a_n) \text{ have the same distribution} \end{aligned} \quad (7)$$

$$\begin{aligned} \forall (a_1, \dots, a_n) \in \mathcal{A}, \forall c \in \mathcal{M}', \\ c(a_1, \dots, a_n) \text{ have the same distribution} \end{aligned} \quad (8)$$

$$\begin{aligned} \forall (a_1, \dots, a_n) \in \mathcal{A}, \forall c \in \mathcal{M}', \\ \mathbb{E}(c(a_1, \dots, a_n)) \text{ are equal} \end{aligned} \quad (9)$$

$$\begin{aligned} \forall (a_1, \dots, a_n) \in \mathcal{A}, \\ (a_1, \dots, a_n) \text{ have the same distribution} \end{aligned} \quad (10)$$

PROOF. Eq. 10 implies Eq. 7 which in turn implies Eq. 8 which in turn implies Eq. 9. We just have to show that Eq. 9 implies Eq. 10. Eq. 9 states that all the (a_1, \dots, a_n) have the same cumulative distribution function, and therefore the same distribution. This implies Eq. 10. \square

For the sake of simplicity we will prefer the last statement in our NFL definitions: two algorithms will be equivalent if $Y(f, m, a)$ and $Y(f, m, b)$ follow the same distribution.

No-Free-Lunch definitions

We introduce here different notions of NFL. We generalize first the definition of No-Free-Lunch in order to extend results stated in Theorem 2.2. To do so, we consider the simple c.u.p. subset $\mathcal{F}_{f_0}^{\Pi}$ introduced in Eq. 5 and satisfying Theorem 2.2. A permutation π on \mathcal{X} is a bijective (or one-to-one) mapping from \mathcal{X} to \mathcal{X} . This definition also holds when \mathcal{X} is uncountable. In some cases, we will have to consider measure-preserving permutations:

DEFINITION 2.4 (MEASURE PRESERVING). Let $(\mathcal{X}, \mathfrak{B}, \mu)$ be a measure space, and $T : \mathcal{X} \rightarrow \mathcal{X}$ be a measurable transformation. We call T measure-preserving if for all $A \in \mathfrak{B}$, we have that $\mu(T^{-1}(A)) = \mu(A)$.

When \mathcal{X} is a discrete space and μ the cardinal measure ($\mu(A) = |A|$), any permutation is measure preserving.

For our first definition of NFL we consider distributions of functions $f_0 \circ \pi$ where π is a “random permutation”. We will see other, more general (*i.e.* weaker), forms of NFL in the sequel, and show that weaker forms do not hold in the continuous case.

The intuitive idea of a random permutation π is a random variable with values in $\mathcal{X}^{\mathcal{X}}$. However, the definition of a random variable involves σ -algebra: a mapping with values in $\mathcal{X}^{\mathcal{X}}$ is a random variable if it is measurable. Therefore, we need a σ -algebra on $\mathcal{X}^{\mathcal{X}}$ for using this terminology. Such σ -algebra exists, but they have many weaknesses [1]. Instead, we will simply require that π is a measurable map as a function $(\omega, x) \in \Omega \times \mathcal{X} \mapsto \pi(x)$ (where, as usual, π implicitly depends on $\omega \in \Omega$).

Our first No-Free-Lunch definition reads as follows:

⁴A characteristic function of a set A is defined to be identically one on A and is zero elsewhere.

DEFINITION 2.5. Let $f_0 : \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function and π a random permutation on \mathcal{X} such that π is measure-preserving with probability one. $\mathcal{NFL}(\mathcal{X}, \pi, f_0)$ holds iff for any integer m (smaller than $|\mathcal{X}|$ when \mathcal{X} is finite) and any two optimization algorithms a and b ,

$$\begin{aligned} Y(f_0 \circ \pi, m, a) \text{ and } Y(f_0 \circ \pi, m, b) \\ \text{follow the same distribution.} \end{aligned} \quad (11)$$

DEFINITION 2.6. Let $f_0 : \mathcal{X} \rightarrow \mathbb{R}$ be a measurable function, $\mathcal{NFL}(\mathcal{X}, f_0)$ holds iff there exists a measure-preserving random permutation π on \mathcal{X} such that $\mathcal{NFL}(\mathcal{X}, \pi, f_0)$ holds.

DEFINITION 2.7. $\mathcal{NFL}(\mathcal{X})$ holds iff $\mathcal{NFL}(\mathcal{X}, f_0)$ holds for all measurable functions $f_0 : \mathcal{X} \rightarrow \mathbb{R}$.

When \mathcal{X} is finite, Theorem 2.2 implies that $\mathcal{NFL}(\mathcal{X})$ holds:

PROPOSITION 2.8. When \mathcal{X} is finite, $\mathcal{NFL}(\mathcal{X})$ holds.

PROOF. Apply Theorem 2.1 to $f_{\mathcal{F}_{f_0}^{\Pi}}$ defined in Eq. 6. \square

We want now to generalize definitions 2.5, 2.6 and 2.7. In particular we do not want to restrict the distribution of fitness functions to $f_0 \circ \pi$ where π is a measure-preserving random permutation. We therefore have to define a general framework for “random variables”, that works also in the case of \mathcal{X} continuous. This is the issue in stochastic processes (also termed random fields when \mathcal{X} is multi-dimensional, e.g. \mathbb{R}^n). A stochastic process f is a mapping $\Omega \times \mathcal{X} : (\omega, x) \mapsto f(x)$ (where, as usual, $f(x)$ implicitly depends on ω). It is termed measurable when it is measurable as a function of both (ω, x) . This condition is not restrictive: it is necessary for making $Y(f, m, a)$ meaningful as a random variable. We implicitly assume in the sequel that all stochastic processes are measurable, otherwise all statements are pointless.

We have no more permutation, but we need the weaker assumption that the median of the fitness values is constant. As we will see in the continuous case, we could relax this hypothesis for proofs of free-lunch if we allow criteria depending on the median of the objective function; also, if we allow translation of the fitness for setting a constant median, there is no more need of this assumption. Our definition of a random fitness distribution is as follows:

DEFINITION 2.9. A random fitness is a measurable map f from Ω to $\mathbb{R}^{\mathcal{X}}$ such that there exists a constant $M_f \in \mathbb{R}$ which is a proper median of f for any ω .

A proper median of a (deterministic) mapping f_0 is a value M_{f_0} such that the measure of $\{x \in \mathcal{X}; f(x) > M_{f_0}\}$ is equal to the measure of $\{x \in \mathcal{X}; f(x) < M_{f_0}\}$ (we assume that \mathcal{X} has a finite measure). Definition 2.9 requires that the proper median exists for any ω , and that it does not depend on ω .

Our generalized definition for No-Free-Lunch reads as follows:

DEFINITION 2.10. Let f be a random fitness. $\mathcal{GNFL}(\mathcal{X}, f)$ holds iff for any $m \in \mathbb{N}$ (smaller than $|\mathcal{X}|$ when \mathcal{X} is finite) and any two optimization algorithms a and b , $Y(f, m, a)$ and $Y(f, m, b)$ follow the same distribution.

This No-Free-Lunch statement depends on the random fitness distribution f . We would like to characterize domains \mathcal{X} for which $\mathcal{GNFL}(\mathcal{X}, f)$ holds for at least one f :

DEFINITION 2.11. $\mathcal{GNFL}(\mathcal{X})$ holds if there is one random fitness f such that $\mathcal{GNFL}(\mathcal{X}, f)$ holds.

Note that we only require the existence of one non-trivial f such that $\mathcal{GNFL}(\mathcal{X}, f)$ holds which is less restrictive than $\mathcal{NFL}(\mathcal{X})$ requiring that $\mathcal{NFL}(\mathcal{X}, f_0)$ holds for any measurable f_0 . The reason is that $\mathcal{GNFL}(\mathcal{X}, f)$ is by nature false for e.g. deterministic non-trivial f . The statement

$$\forall f, \mathcal{NFL}(\mathcal{X}, f) \text{ holds}$$

makes sense, whereas

$$\text{for any random field } f, \mathcal{GNFL}(\mathcal{X}, f)$$

can not hold for (non-trivial) \mathcal{X} .

If there exists f_0 such that $\mathcal{NFL}(\mathcal{X}, f_0)$ holds, the distribution

$$f = \{f_0 \circ \pi, \pi \text{ random permutation s.t. } \mathcal{NFL}(\mathcal{X}, f_0, \pi) \text{ holds}\}$$

is such that $\mathcal{GNFL}(\mathcal{X}, f)$ holds (the median of f is the median of f_0 as π is measure-preserving). Therefore:

PROPOSITION 2.12 (LINK BETWEEN NFLS). For any measure space $(\mathcal{X}, \mathfrak{B}, \mu)$, for any f_0 with a proper median,

$$\begin{aligned} \mathcal{NFL}(\mathcal{X}, f_0) \text{ holds} &\Rightarrow \exists \pi \text{ s.t. } \mathcal{GNFL}(\mathcal{X}, f_0 \circ \pi) \text{ holds} \\ &\Rightarrow \mathcal{GNFL}(\mathcal{X}) \text{ holds} \end{aligned}$$

3. COUNTABLE (NO)-FREE-LUNCH

In this section \mathcal{X} is countably infinite, without loss of generality $\mathcal{X} = \mathbb{N}$. We start by building a non-trivial measurable function f_0 such that $\mathcal{NFL}(\mathcal{X}, f_0)$ does not hold.

PROPOSITION 3.1. If $\mathcal{X} = \mathbb{N}$ and $f_0(i) = i$ for all $i \in \mathbb{N}$, then there is no random permutation π such that $\mathcal{NFL}(\mathcal{X}, \pi, f_0)$ holds. Therefore, $\mathcal{NFL}(\mathcal{X}, f_0)$ does not hold⁵.

PROOF. Assume that such a π exists.

Then, consider, for any $i \in \mathbb{N}$, the algorithm that always chooses $x_1 = i$ as first iterate.

Then, the property $\mathcal{NFL}(\mathcal{X}, \pi, f_0)$, i.e. equation 11, applied to this algorithm for any two different values of i , leads to

$$\mathbb{P}(f_0(\pi(i)) = 1) \text{ is the same for all } i,$$

which, thanks to the definition of f_0 , leads to

$$\mathbb{P}(\pi(i) = 1) \text{ is the same for all } i.$$

But as the events $\{\pi(i) = 1\}$ for $i \in \mathbb{N}$ are a partition, we have

$$1 = \sum_{i \geq 0} \mathbb{P}(\pi(i) = 1) \quad (12)$$

This yields the expected contradiction as $\mathbb{P}(\pi(i) = 1) = 0$ and $\mathbb{P}(\pi(i) = 1) > 0$ both lead to a contradiction (resp. $\sum_{i \geq 0} \mathbb{P}(\pi(i) = 1) = 0$ and $\sum_{i \geq 0} \mathbb{P}(\pi(i) = 1) = \infty$).

⁵Note that any f_0 such that $i \neq j \Rightarrow f(i) \neq f(j)$ would work.

□

This Lemma shows that contrary to finite domain, $\mathcal{NFL}(\mathcal{X})$ does not hold. However, one can study the existence of distributions of functions for which all algorithms are equivalent. In particular if the objective function is a random function with “enough independence” (see proof below), one does not expect any optimization algorithm to be better than another one. This is what we formalize in the following proposition.

PROPOSITION 3.2 (NO FREE COUNTABLE LUNCH).

When $\mathcal{X} = \mathbb{N}$, there exists a distribution f of non-trivial fitness functions (i.e. with minimum different of the maximum with probability one) such that $\mathcal{GNFL}(\mathbb{N}, f)$ holds. Moreover, one can choose f such that there is some f_0 such that with probability 1, $f = f_0 \circ \pi$, with π a permutation of \mathbb{N} . Therefore $\mathcal{NFL}(\mathbb{N}, f_0)$ holds.

PROOF. Let f be the random fitness distribution such that the $f(i)$, for $i \in \mathbb{N}$, are independent and uniformly distributed in $\{0, 1\}$. Technically, such a fitness can be built thanks to Kolmogorov’s extension theorem.

Then, for any algorithm a , and any m ,

$$Y(f, m, a) \text{ is uniformly distributed on } \{0, 1\}^{\{1, \dots, m\}}$$

and therefore $\mathcal{GNFL}(\mathbb{N}, f)$ holds.

Let us now show the second statement of the theorem, i.e. that for some f_0 , $\mathcal{NFL}(\mathbb{N}, f_0)$ holds.

With probability one, the subsets of \mathbb{N} , $\{f^{-1}(1)\}$ and $\{f^{-1}(0)\}$ are infinite. Consider the deterministic fitness function f_0 defined as $f_0(i) = 1$ if i is even and $f_0(i) = 0$ otherwise. Let π be the random permutation defined for all $m \in \mathbb{N}$ as follows:

- if $f(m) = 1$, then $\pi(m) = 2 \times k(m)$ with $k(m)$ minimal such that $2k(m)$ is different from $\pi(i)$ for any $i < m$
- if $f(m) = 0$, then $\pi(m) = 2 \times k(m) + 1$ with $k(m)$ minimal such that $2k(m) + 1$ is different from $\pi(i)$ for any $i < m$.

Then $\mathcal{NFL}(\mathbb{N}, f_0, \pi)$ holds.

□

From this theorem we deduce the following corollary:

COROLLARY 3.3. $\mathcal{NFL}(\mathbb{N}, f_0)$ holds with $f_0(x) = 1$ for x even and $f_0(x) = 0$ for x odd.

We summarize the results of Proposition 2.8, Proposition 3.1 and Proposition 3.3 in the following Theorem:

THEOREM 3.4 (NFL IN DISCRETE SPACES). The following holds:

- If \mathcal{X} is finite, then $\mathcal{NFL}(\mathcal{X})$ holds, and therefore $\mathcal{GNFL}(\mathcal{X})$ holds.
- If \mathcal{X} is countably infinite:
 - $\mathcal{GNFL}(\mathcal{X}, f)$ holds with $f(x) = 1$ with probability $\frac{1}{2}$, independently for each x .
 - Therefore, $\mathcal{GNFL}(\mathcal{X})$ holds.
 - $\mathcal{NFL}(\mathcal{X}, f_0)$ holds with $f_0(x) = 1$ for x even and $f_0(x) = 0$ for x odd.
 - but $\mathcal{NFL}(\mathcal{X})$ does not hold, and for any f_0 such that $i \neq j \Rightarrow f_0(i) \neq f_0(j)$, $\mathcal{NFL}(\mathcal{X}, f_0)$ does not hold.

4. CONTINUOUS FREE-LUNCH

In this Section, we show the main result of this paper, namely that there is no (non-trivial) distribution of fitness for which all algorithms are equivalent in the sense of definition 2.10 when \mathcal{X} is a continuous domain. Without loss of generality we assume that $\mathcal{X} = [0, 1]$ and $\mathcal{Y} = \mathbb{R}$.

It is known that the measurability issue in stochastic processes and random fields is non-trivial ([5]). Consider \mathcal{X} a continuous domain. One can not just set a covariance kernel ($\mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$) on \mathcal{X} , marginal laws, and consider "a" random field f with such marginals and covariances. Even with null covariance, one can not just define marginal laws, and consider "a" random variable with such independent marginals. It is not better with just one distribution of probability on \mathbb{R} : one can not simply define a random field with independent marginals and all marginal distributions equal to the required distribution of probability. This works with countable domains, as shown by Kolmogorov's extension theorem ([4, 9]), but this does not work in continuous domains.

THEOREM 4.1 (CONTINUOUS FREE-LUNCH). *Let f be a random distribution of functions with values in $\mathbb{R}^{[0,1]}$ and a constant proper median M_f . Then $\mathcal{GNFL}([0, 1], f)$ does not hold.*

The proof of this theorem relies on Lemma 4.2 and Lemma 4.5 that are stated and proved below.

Proof of Theorem 4.1: Assume that such a f exists. Then, consider two optimization algorithms for just 1 iterate:

- the algorithm deterministically choosing a given $x \in [0, 1]$;
- the algorithm deterministically choosing a given $x' \in [0, 1]$;

Assuming that $\mathcal{GNFL}([0, 1], f)$ holds leads to the fact that $f(x)$ is distributed as $f(x')$. Therefore, all the $f(x)$ for $x \in [0, 1]$ are identically distributed. Hence, we have shown that in the continuous domain, no-free-lunch theorems imply the identical distribution of the fitness-values of each point in the domain.

Define $g(x) = 1$ is $f(x) > M_f$, where M_f is the median of f , and $g(x) = 0$ otherwise. g is another random field.

First, $\mathcal{GNFL}([0, 1], f)$ implies that the $g(x)$ are identically distributed for $x \in [0, 1]$.

Now, let us show that the $g(x)$ are independent for $x \in [0, 1]$. Consider some fixed x_1, \dots, x_{n-1}, x_n in $[0, 1]$ (all different). $(g(x_1), \dots, g(x_n))$ must be, if $\mathcal{GNFL}([0, 1], f)$ holds, the same as for random search.

For random search, $(g(x_1), \dots, g(x_n))$ is uniformly distributed on $\{0, 1\}^n$.

Therefore, the $g(x_1), \dots, g(x_n)$ are independent. g is a random field, such that $g(x) = 1$ with probability $\frac{1}{2}$ and $g(x) = 0$ otherwise, for any x , and all the $g(x)$ are independent.

We then conclude by Lemma 4.2 that such a g can not exist. □

Kolmogorov's extension theorem ensures that for any distribution of probability, sequences of independent random variables with the same distribution can be built. Some

extensions exist for continuous cases, but without independence. The interested reader is referred to random field theory for this point. We here show that a fully independent and identically distributed family of non-constant random variables indexed by the continuum can not be defined.

LEMMA 4.2 (NO CONTINUOUS FAMILY OF I.I.D. RV). *Assume that g is a random function with values in $\{0, 1\}^{[0,1]}$, such that all the $g(x)$ for $x \in [0, 1]$ are identically distributed. Assume that almost surely g is Lebesgue measurable (so that all probabilities are well defined). Assume that there is $p \in]0, 1[$ such that $p = \mathbb{P}(g(x) = 1)$ for all x . Then, the $g(x)$ are not independent.*

PROOF. Assume that the $g(x)$ are independent. We will now try to get a contradiction. We will split the proof in the following steps:

1. We show (by Lemma 4.5) that almost surely $g^{-1}(1)$ has Lebesgue measure p .
2. We apply Step 1 to show that almost surely, there is an open interval in which $g^{-1}(1)$ has average density $\geq p' > p$.
3. We apply Step 2 to show (by countability) that there is at least one $(a, b) \in \mathbb{Q}^2 \cap [0, 1]^2$ with $a < b$ such that $g^{-1}(1)$ has average density $> p'$ in $[a, b]$ with positive probability.
4. We show (by Lemma 4.5) that for any a, b rationals in $[0, 1]$, almost surely $g^{-1}(1) \cap [a, b]$ has Lebesgue measure $p \times (b - a)$.
5. The contradiction arises as Step 3 shows that $g^{-1}(1) \cap [a, b]$ has measure $p'(b - a)$ and Step 4 shows that $g^{-1}(1) \cap [a, b]$ has measure $p \times (b - a)$.

We now present the detailed steps of the proof.

Step 1. We can then apply Lemma 4.5 since all the $g(x)$ are independent, identically distributed with non zero variance (since each $g(x)$ is a Bernoulli random variable with probability $p \in]0, 1[$ to take the value 1). Lemma 4.5 ensures that with probability 1, $g^{-1}(1)$ has a Lebesgue measure of p . This concludes the first step.

Step 2. By Lebesgue's density theorem, $E = g^{-1}(1)$ has density 1 at almost every point in E . We recall below the definition of density and Lebesgue's density theorem:

DEFINITION 4.3. *The density of a set $A \subset \mathbb{R}^d$ in a ϵ -neighborhood of a point $x \in \mathbb{R}^d$, with $\epsilon > 0$, is*

$$d_\epsilon(x, A) = \frac{\mu(A \cap B(x, \epsilon))}{\mu(B(x, \epsilon))}$$

The density of a set A at a point x is

$$d(x, A) = \lim_{\epsilon \rightarrow 0} d_\epsilon(x, A)$$

THEOREM 4.4 (LEBESGUE'S DENSITY THEOREM). *Consider A a measurable set of \mathbb{R}^d . Then, almost any $x \in A$ verifies $d(x, A) = 1$.*

Therefore, almost surely there is a rational segment ⁶ s such that

$$\mu(s \cap E) \geq \left(\frac{1+p}{2}\right) \mu(s) \tag{13}$$

⁶A rational segment is a closed segment included in $[0, 1]$ with rational bounds and non-zero measure.

This concludes step 2.

Step 3. By step 2, and as the set of rational segments is countable, and as almost surely one of them verifies equation 13, there is at least one rational segment which has non-zero probability of verifying equation 13. Consider then $a, b \in \mathbb{Q}^2$ such that $s = [a, b]$ realizes equation 13 with positive probability. This concludes step 3.

Step 4. Consider $g'(x) = g(a+x \times (b-a))$ defined on $[0, 1]$. Apply Lemma 4.5 to g' . Almost surely, by Lemma 4.5

$$g'^{-1}(1) = \{x \in [0, 1]; a + x(b-a) \in E\}$$

has measure p . This implies that

$$g^{-1}(1) \cap [a, b] \text{ has measure } p(b-a). \quad (14)$$

Step 5. Eq. 14 is a contradiction with Eq. 13. □

LEMMA 4.5. *Assume that g is a stochastic process with values in $\{0, 1\}^{[0,1]}$. Assume that almost surely g is Lebesgue measurable. Assume that all the $g(x)$ are identically distributed, have non-zero variance and are independent. Define $E = g^{-1}(1)$. Let $p = \mathbb{P}(g(x) = 1)$ that does not depend on x by hypothesis above. Then, with probability 1, $\mu(E) = p$, with μ the Lebesgue-measure.*

PROOF. Since g is a stochastic process taking values in $\{0, 1\}^{[0,1]}$, the Lebesgue measure of the preimage of 1 under g is a random variable that we denote $m = \mu(g^{-1}(1)) = \mu(E)$. We are going to prove that the random variable m is constant almost surely and equal to p (in other words equal to p with probability one).

Let \mathcal{E} be the expected value of $\mu(E)$, i.e. $\mathcal{E} = \mathbb{E}(\mu(E))$ and \mathcal{V} its variance, i.e. $\mathcal{V} = \text{Var}(\mu(E))$.

By definition, $x \mapsto g(x)$ is distributed as $x \mapsto g(x/2)$ and $x \mapsto g((1+x)/2)$.

Therefore:

- $\mu(E)$ is distributed as $(m + m')/2$, with m' an independent copy of m .
- $v = \text{Var}(\mu(E)) = (v+v)/4 = v/2$, and therefore $v = 0$.

Therefore, $\mu(E)$ has variance 0.

Consider x a uniform variable in $[0, 1]$, independent of g . Using the Fubini Theorem, we have that $\mathbb{P}(g(x) = 1) = p = \mathbb{E}(\mu(E))$. Therefore, $\mu(E)$ is equal to p with probability 1. □

5. SUMMARY AND CONCLUSION

In this paper we have investigated extensions of NFL results for countably infinite and continuous domains. We show that a consequence of NFL theorems for a finite domain \mathcal{X} is that for any algorithms a and b , for any m , the random vectors $Y(f_0 \circ \pi, m, a) = \langle f_0 \circ \pi(x_1), \dots, f_0 \circ \pi(x_m) \rangle$ and $Y(f_0 \circ \pi, m, b) = \langle f_0 \circ \pi(x_1), \dots, f_0 \circ \pi(x_m) \rangle$ follow the same distribution for any objective function f_0 and π random permutation uniformly distributed among all permutations over \mathcal{X} . We investigate how this property generalizes to countably infinite and continuous domains. For a non-trivial measurable objective function f_0 and π a measure preserving random permutation, we define $\mathcal{NFL}(\mathcal{X}, \pi, f_0)$ as the fact that for any integer m and any two optimization

algorithms a and b , $Y(f_0 \circ \pi, m, a)$ and $Y(f_0 \circ \pi, m, b)$ follow the same distribution. For $\mathcal{X} = \mathbb{N}$ we give non-trivial objective functions f_0 such that it is not possible to find a random permutation π such that $\mathcal{NFL}(\mathbb{N}, \pi, f_0)$ holds. We also prove that there exists non-trivial f_0 and a random permutation over \mathbb{N} such that $\mathcal{NFL}(\mathbb{N}, \pi, f_0)$ holds. We define a weaker form of NFL, \mathcal{GNFL} that does not restrict the distribution of fitness to the form $f_0 \circ \pi$. For a non-trivial random fitness distribution f , with constant proper median, we define $\mathcal{GNFL}(\mathcal{X}, f)$ as the fact that for any integer m and any two optimization algorithms a and b , $Y(f, m, a)$ and $Y(f, m, b)$ follow the same distribution. Since there exists f_0 and π such that $\mathcal{NFL}(\mathbb{N}, \pi, f_0)$ holds, the distribution $f = f_0 \circ \pi$ is such that $\mathcal{GNFL}(\mathbb{N}, f)$ holds. When $\mathcal{X} = [0, 1]$, we show that it is not possible to find non-trivial distribution of fitness f such that $\mathcal{GNFL}([0, 1], f)$ holds. Our conclusions can be summarized in the following table:

Domain \mathcal{X}	Finite	Count. inf.	Continu.
$\exists f_0, \mathcal{NFL}(\mathcal{X}, f_0)$ holds	y	y	n
$\forall f_0, \mathcal{NFL}(\mathcal{X}, f_0)$ holds	y	n	n
$\exists f, \mathcal{GNFL}(\mathcal{X}, f)$ holds	y	y	n

In some sense, NFL-theorems are true for extremely hard finite cases (e.g. [6, 7, 8]), but they are false for multi-objective optimization ([2]) and we show in this paper that they are also moderately true in infinite discrete cases and false in continuous spaces. The deep reason for this fact is that in “bigger” spaces (and continuous spaces are “very” big), random fields (and distributions of fitness functions are non-trivial random fields in the continuous case) necessarily have correlations ([5]).

For our analysis, we have kept a finite horizon perspective and investigated non-asymptotic properties by looking at the distribution of $Y(f, m, a)$ for finite values of m . We could reasonably wonder what would be true when considering asymptotic behaviors. We know that many randomized search heuristics –including the random search– asymptotically find the essential minimum of any fitness function, almost surely whereas some non-repeating algorithms (e.g. any deterministic algorithms) fail for this asymptotic property. Therefore, some algorithms are better than others from the point of view of the asymptotic behavior.

We conjecture that the existence of a median M_f (in definition 2.9) is not necessary for the main result of this paper, i.e. the free-lunch-theorem in the continuous case. However, if we allow a translation of the fitness function, there is no need for such an hypothesis.

Technically speaking, an interesting fact is that in our proofs, measurability plays a positive role and is not only a technical detail that should be used for the mathematical soundness: the continuous case directly relies on measurability. A possible future direction is to take into account the study of random spaces of fitnesses with some separability conditions [5], where separability could be used to characterize “possible” random-fitness-functions. Hierarchies of optimization algorithms might be defined thanks to a proper formalization of continuous optimization problems as separable random fields.

Acknowledgments

The authors would like to warmly thank Marc Schoenauer for his support when dealing with the art of choosing a title.

6. REFERENCES

- [1] P. Billingsley. *Probability and Measure*. John Wiley and Sons, 1986.
- [2] D. Corne and J. Knowles. Some multiobjective optimizers are better than others. In *IEEE Congress on Evolutionary Computation*, 2003.
- [3] J. C. Culberson. On the futility of blind search: an algorithmic view of "No Free Lunch". *Evolutionary Computation*, 6(2):109–127, 1998.
- [4] P. J. Daniell. Integrals in an infinite number of dimensions. *Annals of Mathematics*, 20:281–88, 1919.
- [5] J. Doob. Stochastic process measurability conditions. *Ann. Inst. Fourier, Grenoble*, 25(3 and 4):163–176, 1975.
- [6] S. Droste, T. Jansen, and I. Wegener. Perhaps Not a Free Lunch But At Least a Free Appetizer. In W. Banzhaf, J. Daida, A. E. Eiben, M. H. Garzon, V. Honavar, M. Jakiela, and R. E. Smith, editors, *Proceedings of the First Genetic and Evolutionary Computation Conference (GECCO '99)*, pages 833–839, San Francisco CA, 13–17 1999. Morgan Kaufmann Publishers, Inc.
- [7] S. Droste, T. Jansen, and I. Wegener. Optimization with randomized search heuristics - The (A)NFL theorem, realistic scenarios, and difficult functions. *Theoretical Computer Science*, 287(1):131–144, 2002.
- [8] C. Igel and M. Toussaint. A No-Free-Lunch Theorem for Non-Uniform Distributions of Target Functions. *Journal of Mathematical Modelling and Algorithms*, 3(4):313–322, 2004.
- [9] A. Kolmogorov. *Foundations of the theory of Probability (original: Grundbegriffe der Wahrscheinlichkeitsrechnung)*. Chelsea publishing company, New-York, (1933 original), 1956.
- [10] N. Radcliffe and P. Surry. Fundamental limitations on search algorithms: Evolutionary computing in perspective. In J. van Leeuwen, editor, *Lecture Notes in Computer Science*, Berlin, Germany, 1995. Springer.
- [11] C. Schumacher, M. D. Vose, and L. D. Whitley. The No Free Lunch and Problem Description Length. In L. Spector, E. Goodman, A. Wu, W. Langdon, H.-M. Voigt, M. Gen, S. Sen, M. Dorigo, S. Pezeshk, M. Garzon, and E. Burke, editors, *Genetic and Evolutionary Computation Conference (GECCO 2001)*, pages 565–570, San Francisco, CA, USA, 2001.
- [12] E. VanMarcke. *Random Fields: Analysis and Synthesis*. MIT Press, Cambridge MA, 1998.
- [13] D. Wolpert and W. Macready. No Free Lunch Theorems for optimization. *IEEE Transactions in Evolutionary Computation*, 1(1):67–82, 1997.