A Probabilistic Re-Formulation of No Free Lunch: Continuous Lunches Are Not Free

Alan J. Lockett   alan.lockett@gmail.com
CS Disco, Inc., 4400 Post Oak Pkwy, Suite 2700, Houston, TX, 77027, USA

Risto Miikkulainen   risto@cs.utexas.edu
Department of Computer Science, University of Texas, Austin, 78712, USA

Abstract
No Free Lunch (NFL) theorems have been developed in many settings over the last two decades. Whereas NFL is known to be possible in any domain based on set-theoretic concepts, probabilistic versions of NFL are presently believed to be impossible in continuous domains. This article develops a new formalization of probabilistic NFL that is sufficiently expressive to prove the existence of NFL in large search domains, such as continuous spaces or function spaces. This formulation is arguably more complicated than its set-theoretic variants, mostly as a result of the numerous technical complications within probability theory itself. However, a probabilistic conceptualization of NFL is important because stochastic optimization methods inherently need to be evaluated probabilistically. Thus the present study fills an important gap in the study of performance of stochastic optimizers.

1   Introduction

The No Free Lunch (NFL) theorems were originally proposed in a probabilistic form, stating loosely that on average all non-repeating black-box algorithms perform equally. In recent years, this probabilistic interpretation has come into question due to the difficulty of expressing probabilistic NFL in infinite search domains. The issue became acute with the claim of Auger and Teytaud (2007, 2010) that "continuous lunches are free". They asserted that there is no way of averaging fitness functions that produces the NFL property for a continuous search domain. This assertion was described by Rowe et al. (2009) as a "failure of the probabilistic framework." Duéñez-Guzmán and Vose (2013) concluded that "probability is inadequate to affirm unconstrained NFL results in the general case", and even that "probability is an unfortunate historical artifact".

The goal of this article is to set the record straight with respect to probabilistic NFL. Although we cannot argue with Rowe et al. in their claim that "probabilistic language . . . complicates both the statement and the proof of NFL results" (Rowe et al., 2009), we believe that this complication is necessary. In terms of applications, NFL is a statement about what kinds of performance are

possible in optimization algorithms. From this point of view, NFL for stochastic optimization cannot be divorced from the probabilistic context. Though Rowe et al. (2009) have shown that set-theoretic results can be applied to a probabilistic context, it is not always natural to express general non-NFL performance evaluation results without reference to probability theory.

Much of the confusion stems from the explicit claim of Auger and Teytaud (2010) that "there is no random fitness for which all algorithms are equivalent ...when [the search space] is a continuous domain." In fact, this claim is only possible due to the particularly restrictive conditions that these authors placed on the definition of a random fitness (see Section 2.2). We will construct a large family of NFL evaluation scenarios in continuous settings, demonstrating that the claim of "no continuous NFL" overstates the actual situation unnecessarily.

In sum, this article should dispose of the claim that "probability is inadequate to affirm unconstrained NFL results in the general case." The revised version of probabilistic NFL expounded in this paper provides novel insight into the nature of NFL from a probabilistic and functional analytic point of view.

2 Overview of Results

This section gives an overview of the approach and results of this paper and provides an intuitive sense for what is proven in technical detail below.

2.1 Search and Optimization

In general terms, an iterative search method picks points from a search domain in a sequence with the goal of finding a point with a desired value. The search domain is denoted as a set X, and the set of values is another set Y. The search domain may consist of real numbers, integers, graphs, or anything else. The values set can be any distinct objects: real numbers, integers, or farm animals.

A *search problem* associates values with search points. There are many kinds of problems. One simple problem chooses values based on a single function $u : X \rightarrow Y$, so that if X is the set of positive integers and Y a set of farm animals, then for each integer $x \in X$, $u(x)$ is the associated farm animal. Such a problem might arise in a database recording the animals owned by a farm.

A *search goal* is needed to complete the setting. If the goal is to find the unique identifier of a sheep named Dolly, a search method might use a secondary index based on the name of each animal to locate the correct sheep in logarithmic time, or to say that no such sheep exists. In this case, this search problem with this goal has an efficiently computable solution.

In an *optimization problem*, the value set Y is an ordered set, either totally or partially. The ordering indicates preference, and the goal to find the search point with the least or the greatest value. Many if not most evolutionary algorithms are applied to optimization. Often, the problem is defined by a single fitness (or cost or objective) function $u : X \rightarrow Y$ that provides fixed values for each search point. This setting is called *static optimization*.

A search method proposes a sequence of search points, and a search problem provides the associated values for each point. The method may randomize its choice of search points, choosing the next point according to a probability distribution that depends on the set of prior points and their associated values. Such a method has been called a *randomized search heuristic* (Vose, 1999). The search problem may also be randomized, and with a (deterministic) search goal the search becomes a game-theoretic game, as explored by Lockett (2015).

Typically, randomness within the search problem is substantially constrained. Here, the problem is restricted to a *fitness measure*, defined in Section 3.6. A fitness measure is a probability distribution over all possible fitness functions from the search domain X to the set of fitness values Y. If the same search point is proposed multiple times, it must still receive the same value each time. A more general form of randomness not considered in this article turns up in practice when a fitness function is subject to noise.

2.2 No Free Lunch as Symmetry of Values

NFL looks at the trajectories of search points and fitness values produced by running a search method on a search problem. NFL is usually expressed in a way that is dependent on the search problem but not the search goal. To discuss NFL, one need not make any assumptions about the value set Y other than that it contains distinct points. In particular, NFL does not care whether the set of values is ordered; such questions are not relevant without a search goal. The reason the idea of a search goal was introduced in the last subsection was to point out that contrary to the folklore surrounding NFL, traditional forms of NFL have little to do with any notion of progress in optimization.

Rather, NFL is a symmetry property applying to the sequence of observed values. One might imagine an iterative search method proceeding as follows:

$$
\begin{array}{llllllllllll}
\text{search method} & x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 & x_8 & x_9 & x_{10} & \cdots \\
\text{search problem} & y_1 & y_2 & y_3 & y_4 & y_5 & y_6 & y_7 & y_8 & y_9 & y_{10} & \cdots
\end{array}
\tag{1}
$$

A sequence of such (x, y) pairs is called a *trace* of the search. NFL refers to various situations in which the sequence of $y_1, y_2, y_3, \ldots$ does not depend on the sequence $x_1, x_2, x_3, \ldots$. Note that such a property does not depend on any relative preference among the y_i's.

Such situations arise in various contexts. *Set-theoretic* NFL describes sets of fitness functions such that composition of any function in the set with any permutation of the search domain yields another function inside the set; that is, the set is closed under permutation (*c.u.p.*) (Schumacher, 2000; Rowe et al., 2009). *Closure under permutation* is equivalent to NFL (Igel and Toussaint, 2004), since it guarantees that no matter what sequence of search points is chosen by the search method, the sequence of values that is obtained cannot say which fitness values will be observed next. Set-theoretic NFL applies to search domains X and fitness spaces Y of any cardinality, though proving as much requires index sets with cardinality potentially greater than that of the integers.

Probabilistic NFL places a conditional probability distribution over fitness sequences given search sequences and states that the probability of observing a particular fitness sequence $y_1, y_2, y_3, \ldots$ is independent of the (randomized) search sequence $x_1, x_2, x_3, \ldots$. Given c.u.p. sets, set-theoretic NFL implies probabilistic NFL when the fitness function is chosen uniformly over the set. However, probabilistic NFL may include notions of NFL not included by the set-theoretic version. Within probability theory, it is difficult but not impossible to express probabilistic NFL for spaces beyond countable cardinality due to the existence of *unmeasurable sets*. This subject is taken up in Sections 3 and 4.

In a previous article, Auger and Teytaud (2010) claimed that "continuous lunches are free". This claim has been widely cited, but it overstates the results proven in their article and contradicts Theorems 4.6 and 4.8 below. Suppose that the search domain is the unit interval $X = [0, 1]$ and the value space is the real line, $Y = \mathbb{R}$. To demonstrate an NFL fitness measure, suppose that each fitness value y_i above is chosen from a standard normal distribution independent of the search sequence $x_1, x_2, x_3, \ldots$ and independent also of the prior values $y_1, y_2, y_3, \ldots, y_{i-1}$. That is, the search method chooses a point, and then the search problem randomly assigns a fitness value to that point that has no relationship to its previous choices or to the search point proposed. Thus the probability of the observing certain fitness values is independent of the search points proposed, satisfying the intuitive criterion for probabilistic NFL. This fitness measure has mutually independent and identically distributed coordinates. It is thus path independent by Theorem 4.6 and has the NFL property by Theorem 4.5. Furthermore, the fitness measure in question does exist as a consequence of the Kolmogorov Extension Theorem.

This example of continuous NFL is intuitive and can be approximated on a digital computer. But it does not have a *proper median* and so does not satisfy the conditions imposed on NFL by Auger and Teytaud (2010). A proper median is a real number such that half of the fitness values are strictly above it and half strictly below. Further, this number must be the same across all fitness functions. If complete fitness functions could be drawn from this NFL fitness measure above (which they cannot, since it is a Baire measure), they would be imbalanced with probability one. The requirement of a proper median is too strong, but the proof against continuous NFL depends on it. That continuous NFL must satisfy certain properties does not imply that it cannot exist at all.

The main difference with Auger and Teytaud (2010) regards the optimization process itself. The present article treats optimization as a stochastic process in which only the trace is observed. In probability theory, an event is *measurable* if a probability can be assigned to it. A probability distribution says nothing about events that are not measurable. Such events cannot be observed. Auger and Teytaud (2010) require that the entire fitness function be observable, that is, that one can sample entire fitness functions. This *Lebesgue process* requirement is critical to their proof. In this paper, the optimization method only observes

the fitness values as they are assigned. If the method is run for countably many steps, it observes countably many fitness values in order. The complete fitness function is unobservable.

This article goes further than just showing that continuous NFL does exist in a probabilistic setting. After all, that fact is a logical consequence of the most recent version of set-theoretic NFL (Rowe et al., 2009). Rather, this article demonstrates equivalence between probabilistic NFL and a new concept of *path independence*, a probabilistic analog to c.u.p sets; it states that probabilistic NFL is invariant with respect to permutations in the search domain.

Path independence is not easy to prove, but it can be easily demonstrated for fitness measures that have mutually independent and identically distributed coordinates, such as the one above. Path independent fitness measures always have identically distributed coordinates (Theorem 4.7), but the converse is not true (Theorem 4.10). It is tempting to think that combining mutual independence and identically distributions would yield path independence and hence NFL. But this conclusion is incorrect, as proven for a special case in Theorem 4.11. NFL fitness measures can be deficient in a sense that c.u.p. sets cannot be. They exhibit correlations among search coordinates that do not disturb NFL by choosing probabilistically between two distinct NFL fitness measures.

Further, weighted combinations of NFL fitness measures are also NFL fitness measures (Theorem 4.9). NFL fitness measures without mutually independent coordinates can arise in exactly this way (Theorem 4.11). The following question then suggests itself: Are there NFL fitness measures that cannot be generated from a convex combinations of uniform priors over sets that are c.u.p.? Or is path independence just the closure under vector operations of mutually independent and identically distributed coordinates? These questions are not answered in this paper, but tools helpful to answer them are developed.

2.3 Motivation for the Formalism

This article considers optimization and search as processes generating a pair of randomized trajectories, one through the search domain X and one through the fitness space Y. The optimization method chooses the search trajectory in response to the fitness trajectory, whereas the optimization problem chooses the fitness trajectory in response to the search trajectory. The optimization method tries to control the fitness trajectory, and NFL describes the situation when this trajectory is uncontrollable. Such processes are not confined to computations; they may describe, for example, any input-output sequence mappings. NFL applies to all such systems irrespective of computational aspects.

The formalism below considers an "optimizer" as a map from fitness trajectories to probability distributions over search trajectories. Intuitively, one can imagine observing a fitness trajectory and asking which search trajectory produced it. Usually, one thinks of an optimizer as examining past search points and fitness values in order to decide the next search point iteratively. Formally,

however, this iterative point of view can be obtained from the formalism by extracting conditional probabilities, as briefly discussed in Section 3.5. More complete discussions can be found elsewhere (Lockett, 2015; Lockett and Miikkulainen, 2013; Lockett, 2013). In particular, the notation $\mathcal{G}_i(z, u(z))$ in Equation 4 represents how the optimizer chooses the ith point given a sequence z of previous search points and values determined by a fitness function u. The *trace* or history of an optimization is a sample from the *history process* from Equation 7, which takes values determined by the optimizer and fitness measure. One can examine the history process by sampling its coordinate projections.

The iterative nature of optimization is encoded in a totally ordered index set $\mathcal{I}$. If it is finite, optimization proceeds for a fixed number of steps. If it is countable, optimization runs in discrete steps forever. In either case, the trace can be observed with search points and fitness values at each step. Uncountable index sets were introduced to NFL by Rowe et al. (2009), where they were necessary for NFL proofs. Here, uncountable cardinality enables a way of thinking about optimization that transcends computation, though this perspective will not be fully developed. One can imagine exact gradient descent as an example with $\mathcal{I} = [0, \infty)$ in which a system starts at a fixed point and follows derivatives to the nearest local optimum taking infinitesimal steps. A further possibility, not explored here, would be to separate the search process from its observation, so that an uncountably indexed search might be observed at finite intervals. Such problems occur naturally in robotics, for example.

Probability theory pertains to what can be observed. At the outset, one specifies the events that can be measured. For probabilistic NFL in this paper, the only observable aspect of optimization is its trace (i.e., Equation 7). Complete fitness functions can only be observed if all possible values are contained in the trace. Otherwise, the actual fitness function is a latent, unobservable factor. Other choices are possible; Auger and Teytaud (2010) required the entire fitness function to be observable. The existence of NFL proven here follows from weaker observability requirements, resulting in a technical distinction between *Borel* and *Baire* optimizers. A Baire optimizer can only observe countably many optimization steps, whereas a Borel one can observe uncountably many. Thus a Borel optimizer is needed whenever $\mathcal{I}$ is uncountable.

If the index set $\mathcal{I}$ has countable cardinality, the history process can be approximately sampled one point at a time to generate an approximate trace. The word "approximate" is necessary because very often the spaces being searched consist of real numbers that cannot be represented directly in a digital computer. Not only that, but the probabilities being sampled are also real numbers with infinite representations, so that sampling is only approximate. Many methods claim to optimize real functions though they are computed with a finite subset of the rational numbers. This theory applies to arbitrary spaces. Of course, when a digital computer is used, finite representations are needed, and there are at most countably many programs available to approximate sets of optimizers

that may have uncountable or higher cardinality depending on the space.

In a theoretical setting, it is not always necessary or useful to limit oneself solely to computable objects. A physical robot is one example of an optimization that is only partially digital. For arbitrarily indexed search processes, NFL is precisely the situation in which fitness trajectories decouple from the search trajectories supposedly driving them. This generalized concept of NFL may in fact have repercussions in mathematics well beyond search and optimization.

3 Formal Grounding: Optimization Methods and Problems

As discussed, NFL for optimization can be viewed as a statement about the invariance of trajectories observed in fitness space when running a search method on a search problem. The first step in studying NFL, then, is to formalize what is meant by the terms *search method* and *search problem*. This section develops an abstraction of optimization methods as functionals from the space of fitness functions to a space of probability measures. A similar construction, but restricted to observing countably many search points, was developed by Lockett and Miikkulainen (2013); Lockett (2013).

3.1 Basic Setting

In this study, the domain of optimization (X, τ_X) is a Hausdorff topological space (that is, a space in which consistent open sets can be defined and in which any two distinct points are separated by some open sets). Optimization is performed with respect to a fitness function that maps points in X to some topological space (Y, τ_Y), called the *fitness space*, the *fitness domain*, or the *co-domain*. The space of all objective functions is Y^X, which is also a topological space under the product topology. As discussed above the space Y need not be ordered and represents settings including single-objective optimization, multiobjective optimization, and gradient-based methods. The topological setting in this article is used to induce probability spaces, and thus topological concepts are restricted to general facts about open, closed, and compact sets (see e.g. Munkres (2000)).

3.2 Index Sets

An iterative optimizer looks at points in the search domain in order. To provide a notion of order, a totally ordered index set $\mathcal{I}$ will be used. $\mathcal{I}$ is assumed to have a least element (here called 1, so that indices begin at 1), so that optimization has a starting point. The index set may have any cardinality, whether finite, countable, continuous, or larger, although the main focus here will be on index sets with at most countable cardinality. Traditionally, countably many steps are enough to formalize iterative optimization methods, but it is possible to imagine methods that follow continuous paths as well. Set-theoretic NFL requires arbitrary cardinality in order to exhaust the search domain in NFL proofs (Rowe et al., 2009). Cardinality plays a secondary role in this paper.

For $i, j \in \mathcal{I}$, the intervals of $\mathcal{I}$ are denoted by $[i, j] = \{k \in \mathcal{I} \mid i \leq k \leq j\}$ for

the closed interval, $(i, j) = \{k \in \mathcal{I} \mid i < k < j\}$ for the open interval, and $[i, j) = \{k \in \mathcal{I} \mid i \leq k < j\}$ and $(i, j] = \{k \in \mathcal{I} \mid i < k \leq j\}$ for the half-open intervals. If $j < i$ these intervals are obviously empty. The notation $X^{\mathcal{J}} = \prod_{j \in \mathcal{J}} X$ for an index set $\mathcal{J}$ indicates the product space formed by taking one copy of X for each $j \in \mathcal{J}$. The index set $\mathcal{I}$ may be larger or smaller than X. Sometimes it is necessary to index only unique elements in X. When this is needed, an index set $\mathcal{I}_X$ will be used, defined so that $\mathcal{I}_X = \{i \in \mathcal{I} \mid |[1, i]| \leq |X|\}$.

A *sequence* is a function from some index set to some domain, denoted by parentheses, e.g. $z = (z_j)_{j \in \mathcal{I}}$. For a subset $\mathcal{K} \subseteq \mathcal{I}$, the restriction of z to $\mathcal{K}$ is denoted by $z_{\mathcal{K}} = (z_j)_{j \in \mathcal{K}}$. Restrictions to intervals are given by e.g., $z_{[1,i]}$ or $z_{[i,j)}$ for $i, j \in \mathcal{I}$. A sequence is finite if its index set has finite cardinality, countable if its index set has countable cardinality, and so on. For any function f, $f(z) = (f(z_j))_{j \in \mathcal{J}}$ indicates pointwise application of the function.

For any set $\mathcal{Z}$, a *permutation* $\pi : \mathcal{Z} \to \mathcal{Z}$ is a bijection on the set that re-arranges its elements. If $z = (z_j)_{j \in \mathcal{J}}$ is a sequence, a permutation π on $\mathcal{J}$ yields $(z_{\pi(j)})_{j \in \mathcal{J}}$ as the sequence z with its elements reordered according to π. The notation $\pi(a)$ means $(a_{\pi(j)})_{j \in \mathcal{J}}$. Similarly, for any set B and any function $f : \mathcal{J} \to B$, the notation $f \circ \pi$ is the composition of f and π, $f \circ \pi(i) = f(\pi(i))$.

For any $\mathcal{K} \subseteq \mathcal{J}$ and any topological space T, a *projection* is a function $\phi_{\mathcal{K}} : T^{\mathcal{J}} \to T^{\mathcal{K}}$ given by $\phi_{\mathcal{K}}(z) = z_{\mathcal{K}}$. A projection $\phi_{\mathcal{K}}$ is finite, countable, or uncountable depending on whether $\mathcal{K}$ is finite, countable, or uncountable.

3.3 Probability Spaces

In the NFL theorems below, potential optima are sampled from Kolmogorov probability measures. A *probability space* is a triple $(\Omega, \Sigma, \mathbb{P})$ where $\mathbb{P}$ is a probability measure over the σ-algebra Σ. A σ-algebra on a space Ω is a set of subsets of Ω for which probabilities can be meaningfully defined; it contains the empty set and Ω and is closed under complements, countable unions, and finite intersections. Sets in a σ-algebra are called *events*, *measurable events*, or *measurable sets*. A probability measure $\mathbb{P}$ is a nonnegative set function defined on all sets of a σ-algebra subject to the requirement that $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\bigcup_n E_n) \leq \sum_n \mathbb{P}(E_n)$ for any countable sequence of events $(E_n)_{n \in \mathbb{N}}$. The technical motivation for these definitions is given by e.g. Billingsley (1986). The set of all probability measures over a σ-algebra Σ will be written as $\mathcal{P}(\Sigma)$.

A function f defined from a probability space $(\Omega, \Sigma, \mathbb{P})$ to a measurable space $(S, \mathcal{S})$ is called *measurable* if for every $E \in \mathcal{S}$ the inverse projection $f^{-1}(E) = \{\omega \in \Omega \mid f(\omega) \in E\}$ is contained in Σ.

On any topological space T, the Borel σ-algebra $\mathcal{B}_T$ is the smallest σ-algebra containing the open and closed sets; probability measures defined on this σ-algebra are called *Borel measures*. A function is *Borel measurable* if it is measurable from the Borel σ-algebra on its input space to a σ-algebra on its output space.

This article studies probability measures on product spaces used to sample countable projections of optimization histories. For an index set $\mathcal{J}$ (usually with

$\mathcal{J} \subseteq \mathcal{I}$), the Baire σ-algebra on $T^{\mathcal{J}}$, denoted $\mathcal{B}a_{T^{\mathcal{J}}}$, is the smallest σ-algebra that makes all countable projections into Borel measurable functions. Probability measures on the Baire σ-algebra are called *Baire measures*. A function is called *Baire measurable* if it is measurable from the Baire σ-algebra on its input space to a given σ-algebra on its output space.

If $\mathcal{J}$ is at most countable, then the Baire and Borel σ-algebras are equivalent, i.e., $\mathcal{B}_{T^{\mathcal{J}}} = \mathcal{B}a_{T^{\mathcal{J}}}$. Otherwise, $\mathcal{B}a_{T^{\mathcal{J}}} \subseteq \mathcal{B}_{T^{\mathcal{J}}}$. In particular, the singletons are always contained in the Borel σ-algebra (that is, the singletons are *Borel measurable*), but the Baire σ-algebra may not contain the singletons (that is, the singletons are not *Baire measurable*). On the space Y^X, one cannot sample entire functions from a Baire measure, one can only observe a countable projection of fitness values. Borel measures, however, can be sampled directly.

For this reason, it is worthwhile to mention some facts relating Baire and Borel measures. In many spaces, every Baire measure can be extended to a unique regular Borel measure that agrees with it on all Baire measurable sets. Such spaces are called *Mařík spaces* or simply *Mařík*. It has long been known that every normal and countably paracompact space is Mařík (1957). The following remark gives more familiar examples, considered particularly when $T^{\mathcal{J}} = Y^X$.

Remark 1. *A product space $T^{\mathcal{J}}$ is Mařík in any of the following cases:*

1. *T is finite.*

2. *$\mathcal{J}$ is finite and T is at most countable.*

3. *$\mathcal{J}$ is finite and T is locally compact and σ-compact, e.g. $T = \mathbb{R}^n$ or $T = \mathbb{N}^n$.*

4. *T is a compact space, e.g. $T = \mathbb{R} \cup \{\pm\infty\}$ or $T = [-M, M]$ for $M < \infty$.*

5. *$T^{\mathcal{J}}$ is a metric space, e.g. T is metric and $\mathcal{J}$ is countable.*

If $\mathcal{J}$ has uncountable cardinality, then $T^{\mathcal{J}}$ is not normal unless T is compact. Thus, for instance, if $\mathcal{J} = \mathbb{R}$ and $T = \mathbb{R}$, then $T^{\mathcal{J}} = \mathbb{R}^{\mathbb{R}}$ is not normal and thus not Mařík.

When Y^X is a Mařík space, it is possible to speak of sampling individual fitness functions. It is important to notice that although $\mathbb{R}^{\mathbb{R}}$ is not Mařík, the extended space $(\mathbb{R} \cup \{\pm\infty\})^{\mathbb{R}}$ is Mařík because $\mathbb{R} \cup \{\pm\infty\}$ is compact.

There is a well-defined notion of integration in any probability space based on the Lebesgue integral. This integral is written as $\int_A f(\omega)\, d\mathbb{P}(\omega)$ or just $\int_A f\, d\mathbb{P}$ for any set $A \in \Sigma$ and any Borel measurable real-valued function f. In probability theory, this integral is called the *expectation* of f over $\mathbb{P}$, and the function f is called a *random variable*. A *stochastic process* is a collection of random variables, here indexed by $\mathcal{I}$, which represents time. For further discussion, see Billingsley (1986); Halmos (1974); Chung and Williams (1990); Karatzas and Shreve (1991).

3.4 Kolmogorov Extensions

The Kolmogorov Extension Theorem (Kolmogorov, 1933) is a key tool for building Baire measures on product spaces. This version comes from Aliprantis and

Border (2006) and uses tightness.

Definition 1 (Tightness). *A probability measure $\mathbb{P}$ on a Borel measurable space $(\Omega, \mathcal{B}_\Omega)$ is tight if for all $E \in \mathcal{B}_\Omega$,*

$$\mathbb{P}(E) = \sup \{\mathbb{P}(K) \mid K \text{ is compact in } \Omega\}. \tag{2}$$

That is, $\mathbb{P}(E)$ is the upper bound on the probability mass of all compact subsets.

Tightness and compactness are only required here in order to apply the Kolmogorov Extension Theorem. Every measure is tight in complete separable metric spaces such as $\mathbb{N}$, $\mathbb{R}^n$, $L^2(\mathbb{R}_n)$, and other common spaces.

The next definition regards the consistency of a family of probability measures. Let $\mathcal{J}$ be any set. Although $\mathcal{J}$ will be used as an index set, it need not be ordered. Suppose we have a collection of Borel probability spaces $\{(X_j, \mathcal{B}_{X_j}) \mid j \in \mathcal{J}\}$. A subset $\mathcal{K}$ of $\mathcal{J}$ can be used to define subsets. A *family of finite dimensional distributions* over $\mathcal{J}$ is a collection of probability measures indexed by all finite subsets of $\mathcal{J}$, e.g., $\{\mathbb{P}_\mathcal{K} \mid \mathcal{K} \subseteq \mathcal{J}, \mathcal{K} \text{ finite}\}$. Each $\mathbb{P}_\mathcal{K}$ must be defined on the Borel product σ-algebra $\mathcal{B}_{X^\mathcal{K}}$. Consistency requires overlapping subsequences to agree on probabilities. [1]

Definition 2 (Kolmogorov Consistency). *Suppose $\{\mathbb{P}_\mathcal{K} \mid \mathcal{K} \subseteq \mathcal{J}, \mathcal{K} \text{ finite}\}$ is a family of finite-dimensional distributions for a set $\mathcal{J}$. This family is consistent if for any other finite subset $\mathcal{H}$ of $\mathcal{J}$ and any $A \in \mathcal{B}_{X^\mathcal{K}}$,*

$$\mathbb{P}_{\mathcal{K} \cup \mathcal{H}}(\{a \in X^{\mathcal{K} \cup \mathcal{H}} \mid a_\mathcal{K} \in A\}) = \mathbb{P}_\mathcal{K}(A), \tag{3}$$

which essentially says that members of the family defined on smaller index sets can be obtained by integrating out all extra indices from members defined on larger index sets.

Theorem 3.1 (Kolmogorov Extension Theorem). *Let $\{X_j \mid j \in \mathcal{J}\}$ be a family of Hausdorff topological spaces over a set $\mathcal{J}$, each equipped with their Borel σ-algebras. Let $\{\mathbb{P}_\mathcal{K} \mid \mathcal{K} \subseteq \mathcal{J}, \mathcal{K} \text{ finite}\}$ be a consistent family of finite-dimensional distributions for $\mathcal{J}$. Assume that each $\mathbb{P}_\mathcal{K}$ is tight. Then there is a unique Baire probability measure $\mathbb{P}$ on $X^\mathcal{J}$ that extends each $\mathbb{P}_\mathcal{K}$.*

As mentioned above, in the case of a Mařík space (Remark 1), this probability measure can be extended further to a regular Borel probability measure. The Kolmogorov Extension Theorem is used primarily in Section 4 to construct NFL optimization settings. The only requirements are consistency and tightness.

3.5 Optimization Methods

A *Baire optimizer* is a setwise Baire-measurable function $\mathcal{A} : Y^X \to \mathcal{P}(\mathcal{B}a_{X^\mathcal{I}})$ that takes a fitness function and maps it to a Baire probability measure on the product space $X^\mathcal{I}$. "Setwise Baire-measurable" means that for all Baire events E

---

[1]Many texts give a definition of consistency that uses permutations of multi-indices. Permutations are not needed in Definition 2 because it is based on ordered sets of subindices. A complete discussion of the proof and technical details of Kolmogorov extensions using similar notation can be found online in an unpublished manuscript of K. Border at http://www.hss.caltech.edu/~kcb/Notes/Kolmogorov.pdf.

in $X^{\mathcal{I}}$, the map $u \mapsto \mathcal{A}(u)(E)$ is Baire-measurable from Y^X to $\mathbb{R}$. Intuitively, the optimizer observes a fitness function $u \in Y^X$ and matches it with a sequence of search points $z \in X^{\mathcal{I}}$ chosen stochastically. Similarly, a *Borel optimizer* is a setwise Borel-measurable function $\mathcal{A} : Y^X \to \mathcal{P}(\mathcal{B}_{X^{\mathcal{I}}})$ that yields Borel probability measures, where "setwise Borel-measurable" means that $u \mapsto \mathcal{A}(u)(E)$ is Borel-measurable for every Borel event E. Every Borel optimizer is also a Baire optimizer. The two are equivalent when X and $\mathcal{I}$ are both at most countable.

Most if not all practical optimization methods can be placed inside this formalism by decomposing these formal optimizers into their step-by-step behavior. The process amounts to breaking up the probability measure $\mathcal{A}(u)$ into smaller parts representing small time scales such that the whole optimizer may be specified from its parts. A complete construction of Borel and Baire optimizers is avoided to save space. The construction requires a well-ordered subset of the index set $\mathcal{I}$. If the total order on $\mathcal{I}$ is already a well-ordering (as when $\mathcal{I} = \mathbb{N}$), then this subset may be all of $\mathcal{I}$. In more general settings, the following definition applies.

Definition 3 (Sieve). *A subset $\mathcal{S}$ of $\mathcal{I} \cup \{\top\}$ is a sieve if $\mathcal{S}$ is well-ordered in $\mathcal{I} \cup \{\top\}$ and contains the least element 1, the maximal element $\top$, and all limit points of $\mathcal{I}$. For $i \in \mathcal{S}$, $succ(i)$ denotes the next element of $\mathcal{S}$.*

The element $\top$ is added as a greatest element to $\mathcal{I}$ (independent of $\mathcal{I}$) so that the interval $[1, \top) = \mathcal{I}$. Each index i in the sieve has a next element, written $succ(i)$. A particularly interesting case when $\mathcal{I} \subseteq \mathbb{N}$ is the sieve $\{1 + Ki \mid i \in \mathbb{N} \cup \{0\}\}$, which represents a population-based algorithm with population size K. For $\mathcal{I} = [1, \infty)$, an example sieve might be $\{1 + k\epsilon \mid k \in \mathbb{N} \cup \{0\}\}$, which examines the fitness at intervals of size $\epsilon > 0$. One could consider what happens as ϵ vanishes, making the sieve arbitrarily fine.

Breaking up the index set, one can define objects that transition the state of an optimizer just before time $i \in \mathcal{I}$ to its state just before $succ(i)$. From these, an optimizer is constructed from a sieve $\mathcal{S}$ and a collection of measurable and tight generators on that sieve, defined below. These properties produce a measure $\mathcal{A}(u)$ for each $u \in Y^X$. The function $u \mapsto \mathcal{A}(u)$ is then a Baire optimizer and can be uniquely extended to a Borel optimizer when $X^{\mathcal{I}}$ is Mařík.

Definition 4. *A generator set for a sieve $\mathcal{S}$ is a collection $\mathcal{G} = \{\mathcal{G}_i \mid i \in \mathcal{S}\}$ with $\mathcal{G}_i : X^{[1,i)} \times Y^{[1,i)} \to \mathcal{P}(\mathcal{B}_{X^{[i,succ(i))}})$, where $X^{[1,1)} \times Y^{[1,1)}$ is the one element set $\{\emptyset\}$.*

Definition 5. *A generator set for a sieve $\mathcal{S}$ is measurable if for all $i \in \mathcal{S}$ and $F \in \mathcal{B}_{X^{[i,succ(i))}}$ the map $(z, v) \mapsto \mathcal{G}_i(z, v)(F)$ is Borel measurable from $X^{[1,i)} \times Y^{[1,i)}$ to $\mathbb{R}$.*

Definition 6. *A generator set for a sieve $\mathcal{S}$ is tight if for all $i \in \mathcal{S}$, $z \in X^{[1,i)}$ and $v \in Y^{[1,i)}$, the measure $\mathcal{G}_i(z, v)$ is tight.*

Theorem 3.2. *Every measurable and tight generator set corresponds to a unique Baire optimizer that has the black-box property. If $X^{\mathcal{I}}$ is a Mařík space (see Remark 1), then this Baire optimizer extends to a unique Borel optimizer as well.*

Theorem 3.2 can be used to convert the familiar one-step behavior of black-

box optimization methods into their infinite-time behavior. In the case where $\mathcal{I} = \mathbb{N}$ and $Y = \mathbb{R}$, several examples of such constructions were developed by Lockett and Miikkulainen (2013); Lockett (2013, 2014). Since it is tangential to the NFL theorems below, a complete proof of Theorem 3.2 will not be offered, except to say that it involves a transfinite induction on integrals of the form

$$\mathcal{A}_u^{[1,succ(i))}(E \times F) = \int_E \mathcal{G}_i(z, u(z))(F) \, d\mathcal{A}_u^{[1,i)}(z), \tag{4}$$

for $i \in \mathcal{I}$, $u \in Y^X$, and E, F Borel events in $X^{[1,i)}$ and $X^{[i,succ(i))}$, respectively. Such integrals are uniquely defined because the generator set is measurable, and they preserve tightness. With careful handling of the limit points of $\mathcal{I}$, a tight and consistent family of finite-dimensional distributions can be defined, to which the Kolmogorov Extension Theorem may be applied.

Given a Borel or Baire optimizer, it is also possible to construct generator sets using Radon-Nikodym derivatives. These generator sets take the form above only when the optimizer obeys a black-box property.

3.6 Optimization Problems

In this paper, an optimization problem is a probability measure used to select fitness functions. The space of fitness functions is Y^X, and either Borel or Baire probability measures on this space may be used. A *Baire fitness measure* is defined as a probability measure on $(Y^X, \mathcal{B}a_{Y^X})$ and a *Borel fitness measure* is a probability measure on $(Y^X, \mathcal{B}_{Y^X})$. It is always possible to define Baire fitness measures, and if Y^X is a Mařík space (see Remark 1), then every Baire fitness measure extends uniquely to a Borel fitness measure. Fitness measures (Baire or Borel) will generally be denoted by $\mathbb{F}$ below.

Baire fitness measures must be treated with caution, since only countable projections are Baire measurable, as discussed in Section 3.3. That is, one can only ever observe the behavior of a Baire fitness measure at a countable number of search points. Thus one cannot always determine the true minimum or maximum of "sample" from a Baire fitness measure, nor is it always possible to assess properties such as continuity, differentiability, or integrability. One cannot simply assume that a Baire fitness measure exists that places probability one on continuous functions; the existence of such a measure needs to be proven.

It is much easier to work with Borel fitness measures, since one can directly measure individual functions (these are the singletons in Y^X). However, Baire fitness measures are easier to construct using the Kolmogorov extension theorem. In general, the Borel fitness measures encountered below are obtained by unique extension from a Baire fitness measure in a Mařík space.

The remainder of this paper considers the interaction of an optimizer and a fitness measure. The following text pairs Borel optimizers with Borel fitness measures and Baire optimizers with Baire fitness measures. In each case, different assumptions can be supported about the index set $\mathcal{I}$.

Definition 7 (Borel Setting). *A triple $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ consisting of an optimizer, a fitness measure, and an index set is called a Borel setting if $\mathcal{A}$ is a Borel optimizer, $\mathbb{F}$ is a Borel fitness measure, and $\mathcal{I}$ has the property that for any interval $(i, j) \subseteq \mathcal{I}$ there is a countable increasing sequence $(k_n)_{n \in \mathbb{N}}$ contained in (i, j) such that $\lim_n k_n = j$ (that is, $\mathcal{I}$ is first-countable).*

Definition 8 (Baire Setting). *If $\mathcal{A}$ is a Baire optimizer, $\mathbb{F}$ is a Baire fitness measure, and $\mathcal{I}$ is a countable index set, then $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is a Baire setting.*

In the Borel setting, a Borel product measure $\mathcal{A} \times \mathbb{F}$ is defined on $X^{\mathcal{I}} \times Y^X$. For any $E \in \mathcal{B}_{X^{\mathcal{I}}}$ and $G \in \mathcal{B}_{Y^X}$, this measure must satisfy

$$\mathcal{A} \times \mathbb{F}(E \times G) = \int_G \mathcal{A}(u)(E) \, d\mathbb{F}(u). \tag{5}$$

There is one unique probability measure that satisfies this equation, a consequence that follows from Carathéodory's extension theorem by forming an algebra from the complements, finite unions, and intersections of all possible $E \times G$ and requiring $\mathcal{A} \times \mathbb{F}$ to satisfy the probability axioms on this algebra.

In the Baire setting, $\mathcal{A} \times \mathbb{F}$ may be defined similarly, but resulting in a Baire product measure rather than a Borel one. The measure $\mathcal{A} \times \mathbb{F}$ is now used to derive a stochastic process for the fitness values observed at unique search points.

3.7 The Fitness Process

In order to discuss NFL, a stochastic process called the *fitness process* will be derived corresponding to the sequence of fitness values observed at unique search points. NFL implies that the fitness process is independent of the search points.

Suppose that $z \in X^{\mathcal{I}}$ is a sequence of search points. For each $i \in \mathcal{I}$, define $W_i^z : Y^X \to Y$ so that $W_i^z(u) = u(z_i)$ is the projection of u onto the ith coordinate in z. The function W_i^z is Borel and Baire measurable for all $i \in \mathcal{I}$. A stochastic process W^z can now be defined for each $z \in X^{\mathcal{I}}$ by

$$W^z = \{W_i^z \mid i \in \mathcal{I}\}, \tag{6}$$

where each element of the process is a random variable from $(Y^X, \mathcal{B}a_{Y^X})$ to $(Y, \mathcal{B}_Y)$. Recall that a stochastic process is just a collection of indexed random variables (i.e., measurable functions).

When z is generated by some optimizer $\mathcal{A}$, there are two sources of randomness, an optimizer $\mathcal{A}$ and a fitness measure $\mathbb{F}$. Thus we consider Baire probability measures on the space $X^{\mathcal{I}} \times Y^X$. With this setting, for all $i \in \mathcal{I}$, define $W_i : X^{\mathcal{I}} \times Y^X \to Y$ so that $W_i(z, u) = u(z_i)$ as before. Once again, W_i is both Borel and Baire measurable. To see why, note that W_i is a composition of three maps $(z, u) \mapsto (z_i, u) \mapsto (z_i, u(z_i)) \mapsto u(z_i)$, each of which is a finite projection and thus Baire-to-Baire or Borel-to-Borel measurable. Their composition is therefore both Borel and Baire measurable. Then there is a collection of Baire measurable random variables W given by

$$W = \{W_i \mid i \in \mathcal{I}\}, \tag{7}$$

and W is a stochastic process with values in Y. It will be called the *history process* since it contains the fitness history of an optimization run. Notice that the definition of W does not depend on an optimizer or a fitness measure.

Given a Baire (or Borel) optimizer $\mathcal{A}$ and a Baire (or Borel) fitness measure $\mathbb{F}$, each random variable W_i in the history process induces a Borel measure on Y. This measure determines the fitness value of the ith search point stochastically. This process needs to be filtered for unique search points.

Filtering is accomplished using stopping times, which are defined with respect to filtrations. The *natural filtration* of a process $(W_i)_{i \in \mathcal{I}}$ is an increasing sequence of σ-algebras $(\mathcal{W}_i)_{i \in \mathcal{I}}$ such that each $\mathcal{W}_i$ is the smallest σ-algebra that makes W_j a measurable function for each $j \leq i$. Intuitively, $\mathcal{W}_i$ represents the information that can be observed from the history process up to time i.

A stopping time of the history process is a function $T : X^{\mathcal{I}} \times Y^X \to \mathcal{I} \cup \{\top\}$ such that for all $i \in \mathcal{I}$, the set $\{(z, u) \in X^{\mathcal{I}} \times Y^X \mid T(z, u) \leq i\}$ is a measurable event in the σ-algebra $\mathcal{W}_i$. That is, it may not be possible to identify in advance when a stopping time will stop, but it is always possible to determine whether it has already stopped. The element $\top$ is a special element disjoint from $\mathcal{I}$ and greater than any element in $\mathcal{I}$ included to account for events that never happen.

Next, the unique coordinates are identified by a sequence of stopping times. Let $\# : X^{\mathcal{I}} \times \mathcal{I} \to \mathcal{I}_X$ be a function mapping a history z and an index j to the number of unique elements of X in $z_{[1,j]}$. The set $\mathcal{I}_X$ is used because the number of unique elements is limited by the size of X. If $\#(z, j) = i$, then there are i unique elements in z in the first j indices of z (including index j). Define a sequence of stopping times $(T_i)_{i \in \mathcal{I}_X}$ such that $T_i(z, u) = \inf \{j \in \mathcal{I} \mid \#(z, j) = i\}$. If the set under the infimum is empty, then $T_i(z, u) = \top$ by convention. In either the Borel or the Baire setting, each T_i is a stopping time since $\#(z, j) = i$ can be determined by examining the first j coordinates of z (Theorem 3.3). Furthermore, the unique stopping times are ordered. If $i < j$, then $T_i < T_j$. Every stopping time induces a σ-algebra that includes events prior to the stopping time. For a stopping time T, let $\mathcal{S}_T^j = \{(z, u) \in X^{\mathcal{I}} \times Y^X \mid T(z, u) \leq j\}$ be the set on which T stops by time j. The stopped σ-algebra is given by

$$\mathcal{W}_T = \left\{ A \in \mathcal{B}a_{X^{\mathcal{I}}} \otimes \mathcal{B}a_{Y^X} \mid A \cap \mathcal{S}_T^j \in \mathcal{W}_j \text{ for all } j \in \mathcal{I} \right\} \tag{8}$$

for the Baire setting. In a Borel setting $\mathcal{B}a_{X^{\mathcal{I}}} \otimes \mathcal{B}a_{Y^X}$ is replaced by $\mathcal{B}_{X^{\mathcal{I}}} \otimes \mathcal{B}_{Y^X}$.

Next, it needs to be shown when there is also a random variable

$$W_{T_i} : X^{\mathcal{I}} \times Y^X \to Y \tag{9}$$

that maps the history process W to the first coordinate $W_j(z, u)$ such that $\#(z, j) = i$. The variable W_{T_i} represents the ith unique search point in the search history. Its existence indicates that the trajectory of unique fitness values can be measured, which is required to define NFL. If W_{T_i} exists, it would be $\mathcal{W}_{T_i}$-measurable. However, there are two potential problems. First, the set $\mathcal{S}_{T_i}^j$

must be proven to be $\mathcal{W}_j$-measurable (that is, T_i must be Borel or Baire measurable). Secondly, W_{T_i} must be proven to be measurable. Finally, the event that $T_i = \top$ needs to be considered. Each of these issues is dealt with in turn.

Theorem 3.3. *For all $i \in \mathcal{I}_X$, the stopping time T_i is Borel measurable in the Borel setting and Baire measurable in the Baire setting with $\mathcal{W}_i$ as the target σ-algebra for measurability in either case.*

Proof. To prove that T_i is Borel (or Baire) measurable, it suffices to show that for all j the sets $\{(z, u) \mid T_i(z, u) < j\}$ and $\{(z, u) \mid T_i(z, u) = j\}$ are Borel (or Baire) measurable, since these sets generate the Borel σ-algebra for the order topology on $\mathcal{I}_X$. The set $\{(z, u) \mid T(z, u) = \top\}$ is another special case, but since this set is the complement of $T^{-1}(\mathcal{I}_X)$, its measurability follows automatically when the above sets are proven measurable.

For the first type of set, note that $\{(z, u) \mid T_i(z, u) \leq 1\}$ is either the entire space (if $i = 1$) or the empty set (if $i > 1$) and is therefore Baire and Borel measurable. Assume for (transfinite) induction that $\{(z, u) \mid T_k(z, u) \leq j\}$ is Borel (or Baire) measurable for all $k < i$. Note the identities

$$\{(z, u) \mid T_i(z, u) < j\} = \bigcup_{k < i} \{(z, u) \mid T_k(z, u) \leq j\} = \bigcup_{k_n \uparrow i} \{(z, u) \mid T_{k_n}(z, u) \leq j\}.$$

(10)

In the Borel setting the existence of a countable increasing sequence $k_n \to i$ was used to reduce a potentially uncountable union to a countable one on the far right in Equation 10. In the Baire setting $(k_n)_{n \in \mathbb{N}}$ simply enumerates the interval $[1, j)$. Because this union is a countable union of measurable sets, the set on the far left of Equation 10 is Borel (or Baire) measurable. To complete the induction, it remains to show that $T^{-1}(\{j\}) = \{(z, u) \mid T_i(z, u) = j\}$ is measurable.

Define g_j and $\#_j$ so that $X^{\mathcal{I}} \times Y^X \xrightarrow{g_j} X^{[1,j]} \xrightarrow{\#_j} [1, j]$ where $g_j(z, u) = z_{[1,j]}$ and $\#_j(z)$ counts unique elements for $z \in X^{[1,j]}$. Applying these maps in sequence yields $T_i^{-1}(\{j\}) = (g_j \circ \#_j)^{-1}(\{i\})$ since both indicate the set of sequences indexed by $[1, j]$ that contain i unique elements. The measurability of $T_i^{-1}(\{j\})$ will follow if for all $j \leq i$ the g_j are Borel (or Baire) measurable and the set $(g_j \circ \#_j)^{-1}(\{i\})$ is Borel measurable. The functions g_j are projections with the same cardinality as $[1, j]$. In the Baire setting, $[1, j]$ is at most countable, and so the g_j are Baire measurable. In the Borel setting, all projections are Borel measurable. Finally, $\#_j$ is a continuous function from $X^{[1,j]}$ to $\mathcal{I}$ with the order topology. It is therefore Borel measurable, and since $\{i\}$ is a closed (and thus Borel) set, so is $\#_j^{-1}(\{i\})$. $\square$

Once T_i is measurable, Proposition 3.4 will demonstrate the measurability of W_{T_i} whenever all T_i take on values inside $\mathcal{I}_X$ with probability one.

Definition 9 (Non-Repeating Optimizer). *A Borel or Baire setting $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is non-repeating if $T_i = i$ with $\mathcal{A} \times \mathbb{F}$-probability one for all $i \in \mathcal{I}_X$. The setting is eventually non-repeating if $T_i < \top$ with $\mathcal{A} \times \mathbb{F}$-probability one for all $i \in \mathcal{I}_X$.*

Proposition 3.4. *If a Borel or Baire setting $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is eventually non-repeating, each element of the collection $U = \{W_{T_i} \mid i \in \mathcal{I}_X\}$ is a Y-valued random variable, and U is a stochastic process with natural filtration $(\mathcal{W}_{T_i})_{i \in \mathcal{I}_X}$.*

Proof. The claim is that W_{T_i} is $\mathcal{W}_{T_i}$-measurable for all i with $|[1, i]| \leq |X|$. For every $E \in \mathcal{B}_Y$ and every $k \in \mathcal{I}_X$, define sets

$$
\begin{aligned}
Z(E) = & \quad \left\{ (z, u) \in X^{\mathcal{I}} \times Y^X \quad \middle| \ u(z_{T_i(z,u)}) \in E \right\} \\
Z(k) = & \quad \left\{ (z, u) \in X^{\mathcal{I}} \times Y^X \quad \middle| \ T_i(z, u) \leq k \right\}.
\end{aligned}
\tag{11}
$$

From Equation 8, the claim will hold if $Z(E) \cap Z(k)$ is in $\mathcal{W}_k$. Define the capped stopping time $S_i^k(z, u) = \min\{T_i(z, u), k\}$, which is a stopping time and is measurable because T_i is measurable by Theorem 3.3. But whereas T_i is measurable with respect to the whole process W, S_i^k is $\mathcal{W}_k$ measurable because it only depends on the first k indices of W.

Because T_i lies in $\mathcal{I}_X$ with probability one for all $i \in \mathcal{I}_X$, it suffices to show that the stopped projection $(z, u) \mapsto u(z_{S_i^k(z,u)})$ is Borel or Baire measurable (depending on the setting) as a map to $\mathcal{W}_k$ for any $k \in \mathcal{I}_X$. First, the map $(i, z, u) \mapsto u(z_i)$ is a composition of measurable projections and is therefore measurable to $\mathcal{W}_k$ for every $k \in \mathcal{I}$ (where the first argument has $i \leq k$ and measurability in the first argument is Borel with respect to the order topology). Second, the map $(z, u) \mapsto (S_i^k(z, u), z, u)$ is measurable due to the measurability of S_i^k. The composition of these two maps is $(z, u) \mapsto u(z_{S_i^k(z,u)})$, which must be measurable as the composition of measurable maps.

The index $i \in \mathcal{I}$ was arbitrary, so the collection $U = \{\mathcal{W}_{T_i} \mid i \in \mathcal{I}_X\}$ is a collection of random variables, as desired. $\square$

The stochastic process U is called the *fitness process* of an optimizer $\mathcal{A}$ on a fitness measure $\mathbb{F}$. Let $U_i = W_{T_i}$. The fitness process induces a measure $\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}$ on $Y^{\mathcal{I}_X}$ derived from $\mathcal{A} \times \mathbb{F}$. This measure is Baire or Borel depending on the setting. First define $\Phi_U : X^{\mathcal{I}} \times Y^X \to Y^{\mathcal{I}_X}$ so that $(\Phi_U(z, u))(i) = U_i(z, u)$, and then for each measurable set E

$$
\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}(E) = \mathcal{A} \times \mathbb{F}\left(\Phi_U^{-1}(E)\right).
\tag{12}
$$

This construction is a standard way of obtaining a measure on a product space from a stochastic process. The function Φ_U is measurable whenever U is; such constructions will be used again below. NFL is formulated in the next section as a statement about the fitness process and its induced measure.

4 No Free Lunch Theorems

NFL has been widely studied by many authors (Radcliffe and Surry, 1995; Wolpert and Macready, 1997; Droste et al., 1997; Culberson, 1998; Schumacher et al., 2001; Igel and Toussaint, 2004; Auger and Teytaud, 2007, 2010; Rowe et al.,

2009). For a recent full review, see Igel (2014). In this section, a new version of probabilistic NFL is formulated that exists in continuous and function spaces, and conditions are given that are both necessary and sufficient for this version of NFL. Specifically, a fitness measure with the NFL property must yield probabilities that are independent of search paths.

4.1 NFL Basics

Wolpert and Macready (1995) introduced the concept of NFL as a property of search algorithms, and Wolpert and Macready (1997) extended the concept to optimization algorithms. Radcliffe and Surry (1995) had previously introduced an NFL-type theorem as well.

Several versions of NFL have been studied. The prior sections have laid the groundwork for *probabilistic NFL*, in which NFL is treated as a statement about the probability governing the fitness values observed. The original statements of NFL were probabilistic, but applied only to finite search domains and fitness spaces. Auger and Teytaud (2007, 2010) extended probabilistic NFL to continuous spaces. The following definition is introduced in this paper.

Definition 10 (Probabilistic NFL). *A Borel (or Baire) fitness measure $\mathbb{F}$ is of class P-NFL on an index set $\mathcal{I}$ if the map $\mathcal{A} \to \mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}$ is constant for all optimizers $\mathcal{A}$ such that the setting $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is an eventually non-repeating Borel (or Baire) setting.*

That is, a Borel (or Baire) fitness measure $\mathbb{F}$ is of class P-NFL if its fitness process is identically distributed for every eventually non-repeating Borel (or Baire) optimizer, with limitations on the index set determined by the choice of Borel or Baire measures. As will be demonstrated, a fitness measure is of class P-NFL if the probability over fitness trajectories is constant across search paths.

4.2 Path Independence

NFL implies that information extracted from any series of fitness evaluations is insufficient to suggest search points where more desirable fitness values may be found. This section provides rigor to this intuition.

Definition 11 (Non-Repeating Sequence). *A sequence $z \in \mathcal{I}_X$ is non-repeating if the set $E(z) = \{x \in X \mid \exists i \in \mathcal{I}_X \text{ s.t. } z_i = x\}$ has cardinality $|E(z)| = |\mathcal{I}_X|$.*

That is, a sequence is non-repeating if it does not repeat elements of X until all elements of X have been exhausted. Recall that $\mathcal{I}_X$ limits $\mathcal{I}$ to the size of X.

In Section 3.7, the stochastic process W^z of projections $W_i^z(u) = u(z_i)$ was defined. As in Equation 12, for each z and each Borel (or Baire) fitness measure, this process induces a Borel (or Baire) measure $\mathbb{W}_{\mathbb{F}}^z$ on $Y^{\mathcal{I}_X}$. To obtain this measure, define $\Phi_{W^z} : Y^X \to Y^{\mathcal{I}_X}$ so that $(\Phi_{W^z}(u))(i) = W_i^z(u)$. Then let

$$\mathbb{W}_{\mathbb{F}}^z(E) = \mathbb{F}\left(\Phi_{W^z}^{-1}(E)\right) \tag{13}$$

for each event E of $Y^{\mathcal{I}_X}$. A fitness measure is of class P-NFL if and only if this measure is independent of non-repeating z, called *path independence*.

Definition 12 (Path Independence)**.** *A Borel (or Baire) fitness measure $\mathbb{F}$ is path independent over an index set $\mathcal{I}$ if the map $z \mapsto \mathbb{W}_{\mathbb{F}}^z$ is constant for all non-repeating z. If $\mathbb{F}$ is not path independent, then it is path dependent.*

The term "path independence" captures the intuition that there are no paths through the search space that provide more information about the fitness value of the next unique point. Only the number of fitness values requested matters.

Path independence is obviously related to class P-NFL, since it implies that whatever points an optimizer chooses to evaluate, the distribution over fitness values is the same. In fact, the two are equivalent. This claim is the central result of this article. To prove the claim, path independence is first shown to be necessary for class P-NFL, then sufficient. These results are combined in the NFL Identification Theorem below.

To demonstrate necessity, deterministic optimizers are used, especially deterministic optimizers that are *fitness agnostic*, that is, independent of the fitness values. These concepts are defined using the indicator function. For a set E, the indicator χ_E has $\chi_E(x) = 1$ if $x \in E$, and $\chi_E(x) = 0$ otherwise. An indicator function is measurable whenever E is measurable.

Definition 13 (Deterministic Optimizer)**.** *A Borel (or Baire) optimizer $\mathcal{A}$ is deterministic if there exists a function $d : Y^X \to X^{\mathcal{I}}$ such that $\mathcal{A}(u)(E) = \chi_E(d(u))$ for each Borel (or Baire) event in $X^{\mathcal{I}}$. The function d is called the deterministic core of $\mathcal{A}$.*

Definition 14 (Fitness Agnostic Optimizer)**.** *A Borel (or Baire) optimizer $\mathcal{A}$ is fitness agnostic if the map $u \mapsto \mathcal{A}(u)$ is constant.*

Definition 15 (Invariant Trajectory)**.** *If a Borel (or Baire) optimizer $\mathcal{A}$ is deterministic and fitness agnostic, then its deterministic core d is constant and equal to some sequence $z \in X^{\mathcal{I}}$, called the invariant trajectory of $\mathcal{A}$.*

Proposition 4.1. *If the index set $\mathcal{I}$ satisfies the conditions for a Borel (or Baire) setting, then for any $z \in X^{\mathcal{I}}$ there is a Borel (or Baire) optimizer $\mathcal{A}_z$ that is deterministic and fitness agnostic and has z as its invariant trajectory. Furthermore, for any Borel (or Baire) fitness measure $\mathbb{F}$, the triple $(\mathcal{A}_z, \mathbb{F}, \mathcal{I})$ is a Borel (or Baire) setting.*

Proposition 4.2. *Suppose $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is a Borel (or Baire) setting, and that $\mathcal{A}$ is deterministic and fitness agnostic with non-repeating invariant trajectory z_0. Then the fitness process has $\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}} = \mathbb{W}_{\mathbb{F}}^{z_0}$.*

Proof. The definitions imply that $(\mathcal{A}, \mathbb{F}, \mathcal{I})$ is a non-repeating setting so that the fitness process U and the history process W coincide. For every Borel (or Baire) event $E \subseteq Y^{\mathcal{I}_X}$, the following equalities hold with $\mathcal{A} \times \mathbb{F}$-probability one:

$$
\begin{aligned}
\Phi_U^{-1}(E) \quad &\overset{\mathcal{A}\times\mathbb{F}-a.s.}{=} \quad \{(z_0, u) \,|\, \exists v \in E \text{ s.t. } \forall i \in \mathcal{I}_X \quad v_i = U_i(z_0, u)\} \\
&\overset{\mathcal{A}\times\mathbb{F}-a.s.}{=} \quad \{(z_0, u) \,|\, \exists v \in E \text{ s.t. } \forall i \in \mathcal{I}_X \quad v_i = W_i(z_0, u)\} \\
&= \quad \{z_0\} \times \{u \,|\, \exists v \in E \text{ s.t. } \forall i \in \mathcal{I}_X \quad v_i = W_i^{z_0}(u)\}, \quad (14)
\end{aligned}
$$

where two sets are equal "$\mathcal{A} \times \mathbb{F}$-*a.s.*" if they have equal measure under $\mathcal{A} \times \mathbb{F}$.

The last equality makes it possible to apply Equation 5 to $\{z_0\} \times \Phi_{W^{z_0}}^{-1}(E)$ since

$$\Phi_{W^{z_0}}^{-1}(E) = \{u \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_i^{z_0}(u)\}. \tag{15}$$

This equation implies

$$\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}(E) = \mathcal{A} \times \mathbb{F}\left(\Phi_U^{-1}(E)\right) = \int_{\Phi_{W^{z_0}}^{-1}(E)} \mathcal{A}(u)(\{z_0\}) \, d\mathbb{F}(u) = \mathbb{F}\left(\Phi_{W^{z_0}}^{-1}(E)\right) = \mathbb{W}_{\mathbb{F}}^{z_0}(E). \tag{16}$$

That is, $\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}} = \mathbb{W}_{\mathbb{F}}^{z_0}$. $\qquad\square$

Lemma 4.3. *If a Borel (or Baire) fitness measure $\mathbb{F}$ is of class P-NFL on an index set $\mathcal{I}$, it is path independent over $\mathcal{I}$.*

Proof. Assume $\mathbb{F}$ is of class P-NFL on $\mathcal{I}$ and path dependent. Take $z_1, z_2 \in X^{\mathcal{I}}$ to be distinct non-repeating search histories such $\mathbb{W}_{\mathbb{F}}^{z_1} \neq \mathbb{W}_{\mathbb{F}}^{z_2}$. Such trajectories must exist since $\mathbb{F}$ is path dependent. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be fitness agnostic optimizers with invariant trajectory z_1 and z_2 respectively. Then Proposition 4.2 implies that $\mathbb{Y}_F^{\mathcal{A}_1} \neq \mathbb{Y}_F^{\mathcal{A}_2}$, which contradicts the claim that $\mathbb{F}$ is of class P-NFL. $\qquad\square$

Lemma 4.4. *If a Borel (or Baire) fitness measure $\mathbb{F}$ is path independent over an index set $\mathcal{I}$, then it is of class P-NFL.*

Proof. Suppose E is any Borel (or Baire) event of $Y^{\mathcal{I}_X}$. From Equation 12,

$$\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}(E) = \mathcal{A} \times \mathbb{F}\left(\Phi_V^{-1}(E)\right). \tag{17}$$

The proof centers on obtaining a Borel (or Baire) event $E_Y \subseteq Y^X$ such that $\Phi_V^{-1}(E) = X^{\mathcal{I}} \times E_Y$, for which $\mathcal{A}$ may be averaged out of Equation 17 using Equation 5. To this end, note that

$$\begin{aligned}
\Phi_V^{-1}(E) &= \{(z,u) \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = V_i(z,u)\} \\
&= \{(z,u) \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_{T_i(z,u)}(z,u)\} \\
&= \left\{(z,u) \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_i^{(z_{T_i})_{i \in \mathcal{I}_X}}(z,u)\right\}. \tag{18}
\end{aligned}$$

In this last equation, $\tilde{z} = (z_{T_i})_{i \in \mathcal{I}_X}$ represents the sequence z filtered for uniqueness. Now $\tilde{z}$ is non-repeating, so path independence implies that for any non-repeating sequence $a \in X^{\mathcal{I}_X}$, with $\mathbb{F}$- and therefore $\mathcal{A} \times \mathbb{F}$-probability one,

$$\begin{aligned}
\Phi_V^{-1}(E) \ &\overset{\mathcal{A} \times \mathbb{F} - a.s}{=} \ \{(z,u) \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_i^a(u)\} \\
&= \quad X^{\mathcal{I}} \times \{u \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_i^a(u)\}, \tag{19}
\end{aligned}$$

where in the second line the z has been pulled outside the set since it does not appear on the right. Renaming this final set as

$$E_Y = \{u \mid \exists v \in E \ s.t. \ \forall i \in \mathcal{I}_X \quad v_i = W_i^a(u)\}, \tag{20}$$

Equation 5 can be applied to obtain

$$\mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}(E) = \int_{E_Y} \mathcal{A}(u) \left(X^{\mathcal{I}} \right) \, d\mathbb{F}(u) = \mathbb{F}(E_Y), \tag{21}$$

where the right-hand side is independent of $\mathcal{A}$. Therefore the map $\mathcal{A} \mapsto \mathbb{Y}_{\mathbb{F}}^{\mathcal{A}}$ is constant, and $\mathbb{F}$ is of class P-NFL on $\mathcal{I}$. $\qquad\square$

Theorem 4.5 (NFL Identification Theorem). *A Borel (or Baire) fitness measure $\mathbb{F}$ is of class P-NFL on an index set $\mathcal{I}$ if and only if it is path independent over $\mathcal{I}$.*

Proof. This result combines Lemma 4.3 and Lemma 4.4. $\qquad\square$

The NFL Identification Theorem characterizes probabilistic NFL as a statement about fitness measures. Such a characterization yields a concept equivalent to class P-NFL that does not depend at all on optimizers. If each path is identified with its histogram as done by Igel and Toussaint (2004), then the necessity of path independence is either previously proven or at least strongly suggested for finite spaces by Igel and Toussaint (2004) and for infinite spaces by Rowe et al. (2009). To our knowledge, this article contains the first discussion of path independence as a probabilistic concept, and the first proof that path independence is also a sufficient condition for NFL.

4.3 Construction of NFL Fitness Measures

According to the NFL Identification Theorem, every path independent fitness measure is of class P-NFL. But it is not obvious how to construct path independent fitness measures. In this subsection, it is demonstrated that path independence follows whenever all coordinate projections are identically distributed and mutually independent. This fact was initially proven by English (1996, 2000) for NFL in finite spaces, but it is extended below for arbitrary Hausdorff spaces. Critically, it implies that NFL priors exist in general.

Consider the collection of coordinate projections $\{F_x \mid x \in X\}$ such that $F_x : Y^X \to Y$ is given by $F_x(u) = u(x)$. Each F_x is both Borel and Baire measurable. Given a Borel (or Baire) fitness measure $\mathbb{F}$, each F_x induces a measure $\mathbb{F}_x$ such that for any Borel subset E of Y, $\mathbb{F}_x(E) = \mathbb{F}(\{u \mid u(x) \in E\})$. Furthermore, for any subset D of X, a projection $F_D : Y^X \to Y^D$ is defined by $F_D(u) = (u(x))_{x \in D}$. Then F_D is always Borel measurable and is Baire measurable if D has at most countable cardinality. These projections also induce Borel measures $\mathbb{F}_D$ such that for an Borel subset E of Y^D,

$$\mathbb{F}_D(E) = \mathbb{F}(\{u \mid \exists v \in E \, s.t. \, \forall x \in D, \ v(x) = u(x)\}). \tag{22}$$

When D is finite, $\mathbb{F}_D$ is called a *finite-dimensional distribution* of $\mathbb{F}$. Next, two useful properties of $\mathbb{F}_x$ and $\mathbb{F}_D$ are introduced.

Definition 16 (Identically Distributed Coordinates). *A Borel (or Baire) fitness measure $\mathbb{F}$ has identically distributed coordinates if for all $x \in X$ the map $x \to \mathbb{F}_x$ is constant.*

Definition 17 (Mutually independent Coordinates). *A Borel (or Baire) fitness measure $\mathbb{F}$ has mutually independent coordinates if for any finite $D \subseteq X$ and any collection of Borel sets $\{E_x \mid x \in D\}$ of Y, the measure $\mathbb{F}_D$ factorizes as*

$$\mathbb{F}_D \left(\prod_{x \in D} E_x \right) = \prod_{x \in D} \mathbb{F}_x(E_x). \tag{23}$$

Finite subsets are used in the definition of mutual independence to avoid a definition of uncountable products and because these are sufficient to prove Theorem 4.6 for Baire fitness measures and even most Borel fitness measures.

Fitness measures of class P-NFL can be constructed by guaranteeing tightness along with mutually independent and identically distributed coordinates. There is one limitation at present in that Borel fitness measures of class P-NFL can only be constructed in this way when Y^X is a Mařík space.

Theorem 4.6 (Construction of Path Independence). *If a Baire fitness measure $\mathbb{F}$ has identically distributed and mutually independent coordinates and all of the finite-dimensional distributions of $\mathbb{F}$ are tight, then $\mathbb{F}$ is path independent over $\mathcal{I}$ and therefore of class P-NFL on $\mathcal{I}$. The same holds for a Borel fitness measure if Y^X is a Mařík space.*

Proof. Suppose $\mathbb{F}$ satisfies the conditions. Let D be any finite subset of X. Mutually independent and identically distributed coordinates together imply that for any collection of Borel sets $\{E_x \mid x \in D\}$ of Y,

$$\mathbb{F}_D \left(\prod_{x \in D} E_i \right) = \prod_{x \in D} \mathbb{F}_x(E_x) = \prod_{x \in D} \mathbb{F}_{x_0}(E_x), \tag{24}$$

where x_0 is a fixed element of X independent of D. Importantly, Equation 24 uniquely determines $\mathbb{F}_D$ based on Carathéodory's extension theorem. Then $\{\mathbb{F}_D \mid D \subseteq X, D \text{ finite}\}$ is a consistent family of finite-dimensional distributions. Each $\mathbb{F}_D$ is a coordinate restriction of a tight measure and is thus tight. By the Kolmogorov Extension Theorem, there is a unique Baire measure $\mathbb{F}_X$ on Y^X that extends the $\mathbb{F}_D$, and it must hold that $\mathbb{F}_X = \mathbb{F}$ since $\mathbb{F}_X$ is unique and its finite-dimensional distributions agree with those of $\mathbb{F}$.

Now let $\pi : X \to X$ be any permutation of X. For any subset D of X, define $\pi(D) = \{\pi(x) \mid x \in D\}$. Define an isomorphism from $\psi_\pi : Y^X \to Y^{\pi(X)}$ so that $\psi_\pi(v) = v \circ \pi$, and note that the restriction ψ_π^D of ψ_π to Y^D has range $Y^{\pi(D)}$. For arbitrary subsets E of Y^D, define $D_\pi(E) = \{\psi_\pi^D(v) \mid v \in E\}$. Revisiting Equation 24, notice that for the collection $\{E_x \mid x \in D\}$ above,

$$D_\pi \left(\prod_{x \in D} E_x \right) = \prod_{x \in \pi(D)} E_{\pi^{-1}(x)} = \prod_{x \in D} E_x \quad \text{so} \quad \mathbb{F}_{\pi(D)} \circ D_\pi \left(\prod_{x \in D} E_x \right) = \prod_{x \in D} \mathbb{F}_{x_0}(E_x). \tag{25}$$

Due to uniqueness, it holds in general that $\mathbb{F}_D = \mathbb{F}_{\pi(D)} \circ D_\pi$.

It remains to identify the distribution $\mathbb{W}^z_{\mathbb{F}}$ with $\mathbb{F}_I$ for some $I \subseteq X$. The set I need not be finite; it will in fact have $|I| = |\mathcal{I}_X|$. All such $\mathbb{F}_I$ are uniquely defined as Kolmogorov extensions of some subset of $\{\mathbb{F}_D \mid D \subseteq X, D \text{ finite}\}$. The extension theorem is necessary because I may be uncountable and uncountable projections are not Baire measurable.

Now fix any $I \subseteq X$ such that $|I| = |\mathcal{I}_X|$ (which may be strictly smaller than $|X|$). Choose $z \in X^{\mathcal{I}_X}$ to be any non-repeating sequence such that $z_i \in I$ for all $i \in \mathcal{I}_X$. It should be clear that z exhausts I, i.e., $I = \{z_i \mid i \in \mathcal{I}_X\}$. For any Baire event E of $Y^{\mathcal{I}_X}$, a Baire event $I_z(E)$ of Y^I can be defined such that

$$I_z(E) = \left\{ \tilde{v} \in Y^I \mid \exists v \in E \, s.t. \, \forall x \in I \, \exists i \in \mathcal{I}_X, \text{ with } z_i = x \text{ and } \tilde{v}(x) = v_i \right\}, \quad (26)$$

which pairs each element $v \in Y^{\mathcal{I}_X}$ with an element $\tilde{v} \in Y^I$ that agrees with it. Consequently, $\mathbb{W}^z_{\mathbb{F}} = \mathbb{F}_I \circ I_z$ since

$$
\begin{aligned}
\mathbb{W}^z_{\mathbb{F}}(E) &= \mathbb{F}\left(\{u \mid \exists v \in E \, s.t. \, \forall i \in \mathcal{I}_X, \, v_i = u(z_i)\}\right) \\
&= \mathbb{F}\left(\{u \mid \exists \tilde{v} \in I_z(E) \, s.t. \, \forall x \in I, \, \tilde{v}(x) = u(x)\}\right) = \mathbb{F}_I(I_z(E)). \quad (27)
\end{aligned}
$$

Finally, for any permutation π of X, let $\pi(z) = (\pi(z_i))_{i \in \mathcal{I}_X}$. Observe that

$$I_{\pi(z)} = D_\pi \circ I_z. \quad (28)$$

By Equation 27 and the fact that $\mathbb{F}_I = \mathbb{F}_{\pi(I)} \circ D_\pi$,

$$\mathbb{W}^{\pi(z)}_{\mathbb{F}} = \mathbb{F}_{\pi(I)} \circ I_{\pi(z)} = \mathbb{F}_{\pi(I)} \circ D_\pi \circ I_z = \mathbb{F}_I \circ I_z = \mathbb{W}^z_{\mathbb{F}}. \quad (29)$$

Now for any non-repeating $\tilde{z} \in X^{\mathcal{I}_X}$ there is some permutation π on X such that $\tilde{z} = \pi(z)$, and thus $\mathbb{W}^z_{\mathbb{F}} = \mathbb{W}^{\tilde{z}}_{\mathbb{F}}$. That is, $\mathbb{F}$ is path independent over $\mathcal{I}$ and therefore of class P-NFL on $\mathcal{I}$. To extend this result to tight Borel fitness measures, if Y^X is Mařík then $\mathbb{F}_X$ extends uniquely to $\mathbb{F}$. Then $\mathbb{F}_I$ may be assumed to be a tight Borel measure whenever $\mathbb{F}$ is, and the result follows. $\qquad \square$

Since mutually independent and identically distributed coordinates together imply path independence, one wonders whether these properties are also necessary. In fact, identically distributed coordinates are a necessary consequence of path independence, proven in Theorem 4.7, but mutual independence is not, proven in Theorem 4.11 below.

Theorem 4.7. *Every path independent Borel (or Baire) fitness measure has identically distributed coordinates.*

Proof. Suppose that $\mathbb{F}$ is path independent without identically distributed coordinates. Then there exist $x, y \in X$ and a Borel subset E of Y such that $\mathbb{F}_x(E) \neq \mathbb{F}_y(E)$. Let $z^x, z^y \in X^{\mathcal{I}_X}$ be non-repeating such that $z^x_1 = x$ and $z^y_1 = y$. But then consider $G = \{v \in Y^{\mathcal{I}_X} \mid v_1 \in E\}$, for which $\Phi^{-1}_{W^{z^x}}(G) = \{u \mid u(x) \in E\}$, and

$$\mathbb{W}^{z^x}_{\mathbb{F}}(G) = \mathbb{F}_x(E) \neq \mathbb{F}_y(E) = \mathbb{W}^{z^y}_{\mathbb{F}}(G), \quad (30)$$

which achieves the desired contradiction. $\qquad \square$

4.4 Existence of P-NFL Fitness Measures

Theorem 4.6 provides a simple way to construct Baire fitness measures of class P-NFL for any Hausdorff search space X and fitness space Y so long as there exists at least one tight Borel measure on Y. If in addition Y^X is Mařík, then a Borel fitness measure can also be constructed.

Theorem 4.8 (Existence of P-NFL). *Let $\mathcal{I}$ be an arbitrary index set. Suppose that there exists at least one tight Borel measure $\mathbb{Q}$ on the fitness space Y. Then there exists a Baire fitness measure of class P-NFL on $\mathcal{I}$ for any search domain X. If Y^X is Mařík (Remark 1), there exists a Borel fitness measure of class P-NFL on $\mathcal{I}$.*

Proof. Assume $\mathbb{Q}$ is a tight Borel measure on Y. For each $x \in X$, define $\mathbb{F}_x = \mathbb{Q}$. For any finite subset D of X, define $\mathbb{F}_D$ to be the unique measure such that for any collection $\{E_x \mid x \in D\}$ of Borel subsets of Y,

$$\mathbb{F}_D \left(\prod_{x \in D} E_x \right) = \prod_{x \in D} \mathbb{F}_x(E_x) = \prod_{x \in D} \mathbb{Q}(E_x). \tag{31}$$

Each $\mathbb{F}_D$ is tight since it is a product of tight measures. Thus the collection given by $\{\mathbb{F}_D \mid D \subseteq X, D \text{ finite}\}$ is a consistent family of finite-dimensional distributions for which each member is tight. By the Kolmogorov Extension Theorem, there is a unique Baire fitness measure $\mathbb{F}$ that extends this collection.

$\mathbb{F}$ has mutually independent and identically distributed coordinates. Therefore $\mathbb{F}$ is path independent over $\mathcal{I}$ and of class P-NFL on $\mathcal{I}$ by Theorem 4.6.

If Y^X is Mařík, then $\mathbb{F}$ may be extended to a unique regular Borel fitness measure $\hat{\mathbb{F}}$. The measure $\hat{\mathbb{F}}$ still has mutually independent and identically distributed coordinates since it agrees with $\mathbb{F}$ on all finite projections. Therefore $\hat{\mathbb{F}}$ is path independent over $\mathcal{I}$ and of class P-NFL on $\mathcal{I}$, again by Theorem 4.6. $\square$

According to Theorem 4.8, P-NFL fitness measures exist for a broad class of search domains and fitness spaces. Some of these are listed as examples below. Although the index set $\mathcal{I}$ is arbitrary in Theorem 4.8, the class P-NFL is only defined with respect to Borel settings (where $\mathcal{I}$ must be first-countable, e.g. $\mathbb{R}$) and Baire settings (where $\mathcal{I}$ must be countable). For larger settings the unique stopping times T_i cannot be proven measurable for larger index sets. If the definition of P-NFL were expanded to other settings, Theorem 4.8 should still be valid, since it does not rely on the unique stopping times directly. An example of such a P-NFL fitness measure was given in Section 2.2.

4.5 NFL and Vector Subpaces of Fitness Measures

From Theorem 4.8, P-NFL fitness measures exist. But there is more to conclude from it than mere existence. A Borel (or Baire) fitness measure is defined as a Borel (or Baire) probability measure on the space of fitness functions. Given two real numbers $\alpha, \beta \in [0, \infty)$ and two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on the same σ-algebra Σ, the set function $\mu = \alpha\mathbb{P} - \beta\mathbb{Q}$ is a *finite signed measure* on

Σ. The space of all (countably additive) finite signed measures is a well-known Banach space, denoted $ca(\Sigma)$. Within this space, the set of probability measures $\mathcal{P}(\Sigma)$ is a closed, convex set. In this subsection, the consequences of embedding fitness measures in $ca(\mathcal{B}_{YX})$ (generalized Borel fitness measures) or $ca(\mathcal{B}a_{YX})$ (generalized Baire fitness measures) are considered.

The definitions of path independence, identically distributed coordinates, and mutually independent coordinates extend naturally to finite signed measures using the same definitions. One could also extend class P-NFL in the same way, but doing so requires a generalization of optimizers to finite signed measures. In this paper, path independence is used as a proxy for P-NFL (their equivalence remains valid in the generalized spaces). The following holds.

Theorem 4.9. *For any index set $\mathcal{I}$, the path independent generalized Borel (or Baire) fitness measures form a closed vector subspace of the generalized Borel (or Baire) fitness measures with identically distributed coordinates, which is in turn a closed vector subspace of all generalized Borel (or Baire) fitness measures, $ca(\mathcal{B}_{YX})$ (or $ca(\mathcal{B}a_{YX})$).*

Proof. **Identically distributed coordinates**. Suppose μ and ν are generalized Borel (or Baire) fitness functions with identically distributed coordinates. For $\alpha \in \mathbb{R}$, $\alpha\mu$ has identically distributed coordinates. For $x, y \in X$, it holds that $(\mu + \nu)_x = \mu_x + \nu_x = \mu_y + \nu_y = (\mu + \nu)_y$. If $\mu_n \to \mu$ in the total variation norm and $(\mu_n)_x = (\mu_n)_y$ for all n, then $\mu_x = \mu_y$. So generalized Borel (or Baire) fitness measures with identically distributed coordinates form a closed subspace.

Path independence. Suppose μ and ν are path independent generalized Borel (or Baire) fitness measures. It is obvious that for $\alpha \in \mathbb{R}$, $\alpha\mu$ is path independent. Also, $\mu + \nu$ is path independent since for any non-repeating $z, \tilde{z} \in X^{\mathcal{I}_X}$,

$$\mathbb{W}^z_{\mu+\nu} = \mathbb{W}^z_\mu + \mathbb{W}^z_\nu = \mathbb{W}^{\tilde{z}}_\mu + \mathbb{W}^{\tilde{z}}_\nu = \mathbb{W}^{\tilde{z}}_{\mu+\nu}. \tag{32}$$

Now suppose $\mu_n \to \mu$ in the total variation norm. In this case for all non-repeating $z \in X^{\mathcal{I}_X}$, $\mathbb{W}^z_{\mu_n} \to \mathbb{W}^z_\mu$, and since the sequence on the left is identical for each z, the limit is likewise identical. That is, the path independent generalized fitness measures form a closed vector subspace of the set of all generalized fitness measures. Since every path independent fitness measure has identically distributed coordinates (the proof of Theorem 4.7 does not require $\mathbb{F}$ to be a probability measure), the subspace of path independent generalized fitness measures is a closed subspace of such measures. $\square$

Denote the closed vector subspace of $ca(\mathcal{B}_{YX})$ consisting of generalized Borel fitness measures that have identically distributed coordinates by $\mathcal{ID}^{\mathcal{B}}_{X,Y}$, and let $\mathcal{ID}^{\mathcal{B}a}_{X,Y}$ be defined analogously for generalized Baire fitness measures. Let $\mathcal{PI}^{\mathcal{I},\mathcal{B}}_{X,Y}$ be the closed vector subspace of $\mathcal{ID}^{\mathcal{B}}_{X,Y}$ consisting of generalized Borel fitness measures that are path independent over $\mathcal{I}$, and similarly define $\mathcal{PI}^{\mathcal{B}a,\mathcal{I}}_{X,Y}$. It is of interest to know whether the subspaces are proper. Although it is obvious that $ca(\mathcal{B}_{YX}) \neq \mathcal{ID}^{\mathcal{B}}_{X,Y}$ whenever X and Y each have at least two

elements, it needs to be shown that $\mathcal{ID}^{\mathcal{B}}_{X,Y} \neq \mathcal{PI}^{\mathcal{B},\mathcal{I}}_{X,Y}$. The inequality will be proven for many but not all cases below, using the following definition.

Definition 18 (Equipartitionable). *A tuple (X,Y) is equipartitionable if X and Y can be partitioned into $X_1, X_2 \subseteq X$ and $Y_1, Y_2 \subseteq Y$ with the following properties:*

1. X_1, X_2, Y_1, Y_2 *are all nonempty Borel sets in their respective topologies.*

2. *There exist bijections ϕ from X_1 to X_2 and ψ from Y_1 to Y_2.*

3. *There exists a tight measure $\mathbb{Q}$ on Y.*

4. *For G Borel in Y_1, the measure $\mathbb{Q}$ has $\mathbb{Q}(\psi(G)) = \mathbb{Q}(G)$.*

5. *For G Borel in Y_2, the measure $\mathbb{Q}$ has $\mathbb{Q}(\psi^{-1}(G)) = \mathbb{Q}(G)$.*

If (X,Y) is equipartitionable, the measure $\mathbb{Q}$ has $\mathbb{Q}(Y_1) = \mathbb{Q}(Y_2)$ because $\psi(Y_1) = Y_2$. As usual, $\psi(G) = \{y_2 \in Y_2 \mid \exists y_1 \in Y_1 \; s.t. \; \psi(y_1) = y_2\}$. In particular, $([0,1], \mathbb{R})$, $(\mathbb{N}, \mathbb{R})$, and $(\{0,1\}^n, \mathbb{R})$ are all equipartitionable. Search spaces with an odd finite size, however, are not equipartionable. This wrinkle can probably be removed with a revised proof. When (X,Y) are equipartitionable, then path independent fitness measures are properly contained.

Theorem 4.10. *If the search space contains at least three elements and is equipartitionable together with the fitness space, and if the index set contains at least two elements, then $\mathcal{PI}^{\mathcal{B}a,\mathcal{I}}_{X,Y}$ is a proper subspace of $\mathcal{ID}^{\mathcal{B}a}_{X,Y}$. If in addition Y^X is Mařík (Remark 1), then $\mathcal{PI}^{\mathcal{B},\mathcal{I}}_{X,Y}$ is also a proper subspace of $\mathcal{ID}^{\mathcal{B}}_{X,Y}$. That is, there exists at least one path dependent fitness measure with identically distributed coordinates.*

Proof. Given the equipartition from Definition 18, construct path independent Baire fitness measures $\mathbb{F}_1$ on $Y_1^{X_1}$ and $\mathbb{F}_2$ on $Y_2^{X_2}$ using the restrictions

$$\mathbb{Q}|_{Y_1}(G) = \frac{\mathbb{Q}(G \cap Y_1)}{\mathbb{Q}(Y_1)} \quad \text{and} \quad \mathbb{Q}|_{Y_2}(G) = \frac{\mathbb{Q}(G \cap Y_2)}{\mathbb{Q}(Y_2)} \tag{33}$$

for Theorem 4.8. The fitness measures $\mathbb{F}_1$ and $\mathbb{F}_2$ will now be combined in a way that destroys path independence while retaining identically distributed coordinates by placing probability one on functions that either 1) take fitness values in Y_1 on X_1 and in Y_2 on X_2 or 2) take fitness values in Y_2 on X_1 and in Y_1 on Y_2. If fitness values in Y_1 are observed for search points in X_1, then points in X_2 will yield fitness values in Y_2 with probability one. By strategically crossing X_1 and X_2, a path can obtain different fitness sequences.

Let j, k range over $\{1, 2\}$. To permit generalizations, let $\psi_{11} = \psi_{22}$ and $\phi_{11} = \phi_{22}$ be the identity functions. Let $\psi_{12} = \psi$, $\psi_{21} = \psi^{-1}$, $\phi_{12} = \phi$, and $\phi_{21} = \phi^{-1}$ using ϕ and ψ from Definition 18. For any Baire event E of Y^X,

$$E_{jk} = \left\{ u \in Y_j^{X_j} \;\middle|\; \exists v \in E \; s.t. \; \forall x \in X_j, \; \psi_{jk} \circ u \circ \phi_{kj}(x) = v(x) \right\} \tag{34}$$

defines four images of the set E under different conditions. The composition $\psi_{jk} \circ u \circ \phi_{kj}$ moves search points from X_j to X_k, takes a fitness value in Y_k, and transfers the fitness to Y_j. If $j = k$, these operations do nothing. Define $\mathbb{F}$ by

$$\mathbb{F}(E) = \Big(\; \mathbb{F}_1(E_{11}) \; \mathbb{F}_2(E_{22}) \; \Big) \; \mathbb{Q}(Y_1) + \Big(\; \mathbb{F}_1(E_{12}) \; \mathbb{F}_2(E_{21}) \; \Big) \; \mathbb{Q}(Y_2). \quad (35)$$

The first term of $\mathbb{F}$ says that with probability $\mathbb{Q}(Y_1)$ fitness functions take values in Y_1 on X_1 and in Y_2 on X_2. The second term says that with probability $\mathbb{Q}(Y_2)$ fitness functions take values in Y_2 on X_1 and in Y_1 on X_2.

It needs to be shown that $\mathbb{F}$ has identically distributed coordinates but is path dependent. For any Borel event G in Y and any $x \in X$,

$$\left\{ u \in Y^X \,\middle|\, u(x) \in G \right\}_{jk} = \begin{cases} \left\{ u \in Y_j^{X_j} \,\middle|\, \psi_{jk} \circ u \circ \phi_{kj}(x) \in G \cap Y_j \right\} & \text{if } x \in X_k \\ Y_j^{X_j} & \text{otherwise.} \end{cases} \quad (36)$$

Thus if $x \in X_1$, , the coordinate projection has

$$\begin{aligned} \mathbb{F}_x(G) &= \mathbb{F}_1 \left(\left\{ u \in Y_1^{X_1} \,\middle|\, u(x) \in G \cap Y_1 \right\} \right) \; \mathbb{Q}(Y_1) \\ &\quad + \mathbb{F}_2 \left(\left\{ u \in Y_2^{X_2} \,\middle|\, \psi^{-1} \circ u \circ \phi(x) \in G \cap Y_2 \right\} \right) \; \mathbb{Q}(Y_2) \\ &= \mathbb{Q}(G \cap Y_1) + \mathbb{Q}(\psi(G \cap Y_2)) = \mathbb{Q}(G), \quad (37) \end{aligned}$$

using the fact that ψ preserves $\mathbb{Q}$-measure and that each $\mathbb{F}_j$ has coordinates identically distributed according to $\mathbb{Q}|_{Y_j}$. Similarly, if $x \in X_2$, then $\mathbb{F}_x(G) = \mathbb{Q}(\psi^{-1}(G \cap Y_1)) + \mathbb{Q}(G \cap Y_2) = \mathbb{Q}(G)$. $\mathbb{F}$ has identically distributed coordinates.

Now choose $x_1^1, x_1^2 \in X_1$ and $x_2 \in X_2$ and let $z^1, z^2 \in X^{\mathcal{I}_X}$ be non-repeating so that $z_1^1 = x_1^1$, $z_2^1 = x_2$, $z_1^2 = x_1^1$ and $z_2^2 = x_1^2$. In this way, z^1 first checks the fitness domain for X_1, followed by X_2. The sequence z^2 starts in X_1 and then checks another point in X_1. It is possible to choose such points because X has at least three elements. Now

$$\mathbb{F} \left(\left\{ u \,\middle|\, W_1^{z^1}(u) \in Y_1, \, W_2^{z^1}(u) \in Y_2 \right\} \right) = \mathbb{F} \left(\left\{ u \,\middle|\, u(x_1^1) \in Y_1, \, u(x_2) \in Y_2 \right\} \right). \quad (38)$$

To expand this equation, set $x_2^2 = x_2$ and define Y_{jk} by

$$Y_{jk} = \left\{ u \,\middle|\, u(x_1^1) \in Y_1, \, u(x_2) \in Y_2 \right\}_{jk} = \left\{ u \in Y_j^{X_j} \,\middle|\, \psi_{jk} \circ u \circ \phi_{kj}(x_j^j) \in Y_k \right\}. \quad (39)$$

Observe that on the one hand, Y_{jk} is empty if $j \neq k$ so that

$$\mathbb{F}_1(Y_{12}) = \frac{\mathbb{Q}\left(\psi(Y_1 \cap Y_2) \right)}{\mathbb{Q}(Y_1)} = 0 \quad \text{and} \quad \mathbb{F}_2(Y_{21}) = \frac{\mathbb{Q}\left(\psi^{-1}(Y_2 \cap Y_1) \right)}{\mathbb{Q}(Y_2)} = 0, \quad (40)$$

and on the other hand,

$$\mathbb{F}_1(Y_{11}) = \frac{\mathbb{Q}\left(Y_1 \cap Y_1 \right)}{\mathbb{Q}(Y_1)} = 1 \quad \text{and} \quad \mathbb{F}_2(Y_{22}) = \frac{\mathbb{Q}\left(Y_2 \cap Y_2 \right)}{\mathbb{Q}(Y_2)} = 1. \quad (41)$$

Consequently,

$$\mathbb{F}\left(\left\{u \,\middle|\, W_1^{z^1}(u) \in Y_1, \, W_2^{z^1}(u) \in Y_2\right\}\right) = \mathbb{Q}(Y_1) = \frac{1}{2}. \tag{42}$$

By contrast, functions that take values in Y_1 on $x_1^1 \in X_1$ take values in Y_1 on $x_1^2 \in X_1$. Thus, with probability one, $W_2^{z^2}(u) = u(x_1^2) \notin Y_2$, and so

$$\mathbb{F}\left(\left\{u \,\middle|\, W_1^{z^2}(u) \in Y_1, \, W_2^{z^2}(u) \in Y_2\right\}\right) = 0. \tag{43}$$

That is, $\mathbb{F}$ is path dependent, and the path independent Borel fitness measures are properly contained. By construction, $\mathbb{F}$ is a Baire fitness measure. If Y^X is Mařík, it can be uniquely extended to a Borel fitness measure. $\qquad\square$

So in important cases, not all fitness measures with identically distributed coordinates are path independent. There also exist Baire fitness measures of class P-NFL that do not have mutually independent coordinates.

Theorem 4.11. *If the search domain has at least two elements, the set of generalized Baire fitness measures with mutually independent coordinates is not closed under vector operations. Furthermore, there exists a Baire fitness measure that is path independent (and hence of class P-NFL) but does not have mutually independent coordinates. If Y^X is Mařík, both statements are true of Borel fitness measures as well.*

Proof. Let E be any Borel event on Y with $E \neq Y$. Choose $\alpha, \beta \in (0, 1)$ such that

$$2\alpha\beta \neq \alpha^2 + \beta^2. \tag{44}$$

Let $\mathbb{Q}^1$ and $\mathbb{Q}^2$ be any tight measures with $\mathbb{Q}^1(E) = \alpha$, and $\mathbb{Q}^2(E) = \beta$.

Pick $x, y \in X$, and construct $\mathbb{F}_1$ and $\mathbb{F}_2$ by Theorem 4.6 to be path independent Baire fitness measures with mutually independent coordinates distributed identically according to $\mathbb{Q}^1$ and $\mathbb{Q}^2$, respectively. Then $\mathbb{F} = \frac{1}{2}\mathbb{F}^1 + \frac{1}{2}\mathbb{F}^2$ is a Baire fitness measure. Set $D = \{x, y\}$. Then

$$\mathbb{F}_D(E \times E) = \frac{1}{2}\mathbb{F}_x^1(E)\,\mathbb{F}_y^1(E) + \frac{1}{2}\mathbb{F}_x^2(E)\,\mathbb{F}_y^2(E) = \frac{1}{2}\left(\alpha^2 + \beta^2\right) \tag{45}$$

using the notation $\mathbb{F}_D$ and $\mathbb{F}_x$ from Definition 17. To have mutual independence,

$$\mathbb{F}_D(E \times E) = \left[\frac{1}{2}(\alpha + \beta)\right]^2 = \frac{1}{4}\left(\alpha^2 + 2\alpha\beta + \beta^2\right), \tag{46}$$

which together with Equation 45 violates Equation 44. That is, $\mathbb{F}$ does not have mutually independent coordinates, and so mutual independence is not closed under vector operations. However, path independence is closed under vector operations by Theorem 4.9, so $\mathbb{F}$ is path independent. Thus $\mathbb{F}$ is path independent without mutually independent coordinates. By Theorem 4.5, it is also of class P-NFL. If Y^X is Mařík, then $\mathbb{F}^1$ and $\mathbb{F}^2$ can be extended to Borel measures without losing path independence, and so in a Mařík space all of the above statements apply to Borel fitness measures as well. $\qquad\square$

The remaining sections discuss these results.

5   Discussion

Many of the conclusions of this paper have been previously known in either qualitative or quantitive terms. The new features of this paper include the following: 1) a proof that probabilistic NFL does exist in continuous domains; 2) a probabilistic construction of NFL in arbitrary Hausdorff search domains and fitness domains, including potentially uncountable index sets; and 3) the introduction of path independence as a concept equivalent to probabilistic NFL. The relationship to prior NFL results was discussed in Section 2. This section offers comments about the significance of the NFL results of this paper.

NFL is the answer to the question of whether a search problem can be structured to make search progress impossible. With NFL, the past is irrelevant to the future, and there is no means of predicting the outcome of any action (Bousquet et al., 2004). Yet NFL-like properties rarely show up in optimization problems drawn from real-world tasks.

This concept was studied by Igel and Toussaint (2004), who concluded that in the set-theoretic case, the absolute number of permutation-invariant sets vanishes doubly exponentially fast as the size of the search and the fitness domain grows. Not only that, but in problems drawn from real-world tasks, there are often locality constraints (i.e. neighboring points have similar fitness) that can never be permutation invariant. If locality constraints hold, they often hold in most places, so the addition of "Almost NFL", where most search paths produce the same or similar fitness trajectories, does not change the situation (Droste et al., 1997). Because physical regularities generally exist, the prior governing real-world problems is likely not of class P-NFL. Consequently, there exist general-purpose optimizers that outperform others on all tasks averaged according to their real-world likelihood.

A fitness measure that is not of class P-NFL need not be particularly specialized. Jiang and Chen (2011) previously showed that a discrete Lipschitz class does not exhibit the NFL property. In general metric spaces, Lipschitz or Hölder growth properties are essentially locality constraints that prevent NFL. Gaussian processes form another example of general non-NFL fitness measures, which are also subject to locality constraints.

If there is a fitness measure $\mathbb{F}$ that is not of class P-NFL, then one can examine the set of optimal optimization methods for a given performance metric. This question was studied by Auger and Teytaud (2010) for continuous domains and later in Lockett (2014) for arbitrary Hausdorff spaces by solving Bellman-type equations. As it turns out, the optimal method can be expressed mathematically, but this expression contains a nested optimization. This recursion suggests that optimization is computationally hard in the general case, with or without NFL results. In some cases the inner optimization is analytically solvable, though some tractable cases were studied by Powell and Ryzhov (2012). In any case, there is a tradeoff in the time spent to generate search points versus the quality of the search points generated, all independent of NFL.

NFL provides a first step towards a categorization of the space of possible optimization problems. It is difficult to see how a purely set-theoretic approach to NFL can be expanded to study optimal behavior of optimizers on non-NFL classes of fitness functions. Thus if one goal for studying NFL is to develop analytic tools that make it possible to study general classes of optimization methods and problems, then it seems difficult to avoid probabilistic language.

6 Conclusion

This article has provided a renovated version of probabilistic NFL to correct misconceptions in the literature, particularly the claim that probabilistic approaches cannot provide an account of NFL in continuous domains. The methods and techniques in this paper require substantial technical development, yet the tools that result are not limited to the study of NFL. It is hoped that these techniques will be used in the future to obtain a variety of additional results about the nature and performance of optimization methods.

Acknowledgements

References

Aliprantis, C. C. and Border, K. D. (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide, 3rd Edition*. Springer, New York, New York.

Auger, A. and Teytaud, O. (2007). Continuous lunches are free! In *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO-2007)*, pages 916–922, New York. ACM Press.

Auger, A. and Teytaud, O. (2010). Continuous lunches are free plus the design of optimal optimization algorithms. *Algorithmica*, 57(1):121–146.

Billingsley, P. (1986). *Probability and Measure*. John Wiley.

Bousquet, O., Boucheron, S., and Lugosi, G. (2004). Introduction to statistical learning theory. In Bousquet, O., von Luxburg, U., and Rätsch, G., editors, *Advanced Lectures in Machine Learning, LNAI*. Springer-Verlag.

Chung, K. L. and Williams, R. (1990). *Introduction to Stochastic Integration*. Birkhauser, Boston, MA.

Culberson, J. C. (1998). On the futility of blind search: an algorithmic view of "no free lunch". *Evolutionary Computation*, 6(2):109–127.

Droste, S., Jansen, T., and Wegener, I. (1997). Optimization with randomized search heuristics – the (a)nfl theorem, realistic scenarios, and difficult functions. *Theoretical Computer Science*, 287:2002.

Duéñez-Guzmán, E. A. and Vose, M. (2013). No free lunch and benchmarks. *Evolutionary Computation*, 21(2).

English, T. (1996). Evaluation of evolutionary and genetic optimizers: No free lunch. In *Proceedings of the Fifth Annual Conference on Evolutionary Programming (EP V)*, pages 163–169.

English, T. (2000). Optimization is easy and learning is hard in the typical function. In *Evolutionary Computation, 2000. Proceedings of the 2000 Congress on*, volume 2, pages 924–931.

Halmos, P. (1974). *Measure Theory*. Springer-Verlag, New York, NY.

Igel, C. (2014). No free lunch theorems: Limitations and perspectives of metaheuristics. In Borenstein, Y. and Moraglio, A., editors, *Theory and Principled Methods for the Design of Metaheuristics*. Springer-Verlag.

Igel, C. and Toussaint, M. (2004). A no-free-lunch theorem for non-uniform distributions of target functions. *Journal of Mathematical Modelling and Algorithms*, 3(4):313–322.

Jiang, P. and Chen, Y. (2011). Free lunches on the discrete lipschitz class. *Theoretical Computer Science*, 412(17):1614 – 1628.

Karatzas, I. and Shreve, S. (1991). *Brownian Motion and Stochastic Calculus*. Springer-Verlag, New York, NY.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*.

Lockett, A. (2013). Measure-theoretic analysis of performance in evolutionary algorithms. In *Proceedings of the 2013 IEEE Conference on Evolutionary Computation (CEC-2013)*. IEEE Pres.

Lockett, A. (2014). Model-optimal optimization by solving bellman equations. In *Proceedings of the 2014 Genetic and Evolutionary Computation Conference (GECCO-2014)*. ACM Press.

Lockett, A. (2015). Insights from adversarial fitness functions. In *Proceedings of the 13th Workshop on Foundations of Genetic Algorithms (FOGA-2015)*. ACM.

Lockett, A. and Miikkulainen, R. (2013). A measure-theoretic analysis of stochastic optimization. In *Proceedings of the 12th Workshop on Foundations of Genetic Algorithms (FOGA-2013)*. ACM.

Mařík, J. (1957). The baire and borel measure. *Czechoslovak Math. J.*, 7.

Munkres, J. R. (2000). *Topology*. Prentice Hall, Upper Saddle River, NJ.

Powell, W. B. and Ryzhov, I. O. (2012). *Optimal Learning*. John Wiley and Sons.

Radcliffe, N. and Surry, P. D. (1995). Fundamental limitations on search algorithms: Evolutionary computing in perspective. In *Lecture Notes in Computer Science 1000*, pages 275–291. Springer-Verlag.

Rowe, J. E., Vose, M. D., and Wright, A. H. (2009). Reinterpreting no free lunch. *Evolutionary Computation*, 17(1):117–129.

Schumacher, C., Vose, M. D., and Whitley, L. D. (2001). The no free lunch and problem description length. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001*, pages 565–570. Morgan Kaufmann.

Schumacher, C. W. (2000). *Black box search: framework and methods*. PhD thesis. AAI9996384.

Vose, M. D. (1999). Random heuristic search. *Theoretical Computer Science*, 229:103–142.

Wolpert, D. H. and Macready, W. G. (1995). No free lunch theorems for search. Technical Report Technical Report SFI-TR-95-02-010, Santa Fe Institute.

Wolpert, D. H. and Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1):67–82.