No Regret Bound for Extreme Bandits

Robert Nishihara
rkn@eecs.berkeley.edu

David Lopez-Paz
david@lopezpaz.org

Léon Bottou
leon@bottou.org

August 13, 2015

Abstract

Algorithms for hyperparameter optimization abound, all of which work well under different and often unverifiable assumptions. Motivated by the general challenge of sequentially choosing which algorithm to use, we study the more specific task of choosing among distributions to use for random hyperparameter optimization. This work is naturally framed in the extreme bandit setting, which deals with sequentially choosing which distribution from a collection to sample in order to minimize (maximize) the single best cost (reward). Whereas the distributions in the standard bandit setting are primarily characterized by their means, a number of subtleties arise when we care about the minimal cost as opposed to the average cost. For example, there may not be a well-defined "best" distribution as there is in the standard bandit setting. The best distribution depends on the rewards that have been obtained and on the remaining time horizon. Whereas in the standard bandit setting, it is sensible to compare policies with an oracle which plays the single best arm, in the extreme bandit setting, there are multiple sensible oracle models. We define a sensible notion of regret in the extreme bandit setting, which turns out to be more subtle than in the standard bandit setting. We then prove that no policy can asymptotically achieve no regret. Furthermore, we show that in the worst case, no policy can be guaranteed to perform better than the policy of choosing each distribution equally often.

1 Introduction

Our motivation comes from hyperparameter optimization and more generally from the challenge of minimizing a black-box objective $f \colon \Omega \to [0,1]$ which we can only evaluate pointwise. As an example, $\omega \in \Omega$ may parameterize the architecture of a convolutional network, and $f(\omega)$ may be the validation error when the network with that architecture is trained on a particular data set. A number of approaches have been applied to the optimization of f including Bayesian optimization, covariance matrix adaptation, random search, and a variety of other methods (for an incomplete list, see [1, 2, 11, 8, 14, 9, 10, 7]).

In some sense, random search is the benchmark of choice. Whereas other approaches work well under various and often unverifiable conditions (such as smoothness or convexity of the objective), random search has strong finite-sample guarantees that hold without any assumptions on the function under consideration. This guarantee is illustrated by the so-called *rule of 59*,[1] which states that the best of 59 random samples will be in the best 5 percent of all samples with probability at least 0.95. More generally, any distribution over the set of hyperparameters Ω induces a distribution μ over the validation error in $[0,1]$. Let F_μ be the cumulative distribution function of μ, and suppose that F_μ is continuous. Suppose that $x_1, \ldots, x_T$ are independent and identically-distributed samples from μ (obtained, for instance, by independently sampling hyperparameters ω_t and evaluating $x_t = f(\omega_t)$ for $1 \le t \le T$). The following is known.

[1]Though they are known, the rule of 59 and Lemma 1 do not appear in Bergstra and Bengio [1], and they are difficult to find in the literature.

Lemma 1. *The distribution of the extreme cost* $\min\{x_1, \ldots, x_T\}$ *is easily described with quantiles. We have* $P(F_\mu(\min\{x_1, \ldots, x_T\}) \leq \alpha) = 1 - (1 - \alpha)^T$. *More specifically,* $F_\mu(\min\{x_1, \ldots, x_T\})$ *is a Beta*$(1, T)$ *random variable.*

Proof. The event $F_\mu(\min\{x_1, \ldots, x_T\}) > \alpha$ happens if and only if $F_\mu(x_t) > \alpha$ for each t, which happens with probability $(1 - \alpha)^T$. Differentiating the resulting cumulative distribution function gives the density function of a Beta$(1, T)$ random variable. □

The generality of Lemma 1 comes at a price. The guarantee is given with respect to the distribution μ, but there is no guarantee about μ itself. Different induced distributions μ may arise from different parameterizations of the hyperparameter space Ω (for example, from the decision to put a uniform or a log-uniform distribution over a coordinate of ω), and the allocation of mass over $[0, 1]$ may vary wildly based on these choices.

Furthermore, the flip side of making no assumptions on the underlying objective is that random search fails to adapt to easy problems. When the objective under consideration satisfies various regularity conditions (as real-world objectives often do), more heavily-engineered approaches will likely outperform random search. That said, it is not clear how to know that a given algorithm is outperforming random search without also running random search. For this reason, the benefits of a potentially faster algorithm are blunted when one must also run the slow algorithm to verify the performance of the fast algorithm.

Given the variety of existing hyperparameter optimization algorithms, it would be desirable to devise a strategy for sequentially choosing which algorithm to use in a way that performs nearly as well as if we had only used the single best algorithm. We consider the simpler problem of choosing which of several distributions over hyperparameters to use for random search. In Theorem 12, we show that even in this simplified setting, no strategy guarantees performance that is asymptotically as good as the single best distribution, at least not without stronger assumptions.

We will frame our negative result in the extreme bandit setting [5] (also called the max K-armed bandit setting [6]). Prior work has focused on designing algorithms that perform asymptotically as well as the single best distribution under parametric (or semiparametric) assumptions on the possible distributions [6, 5]. Instead, we focus on probing the difficulty of the problem, pointing out a number of subtleties that arise in this setting that do not show up in the conventional bandit setting.

2 The Extreme Bandit Setting

Cicirello and Smith [6] introduce the extreme bandit problem (they call it the max K-armed bandit problem) as follows. We are given a tuple of unknown distributions (arms) $\mu_1^K = (\mu_1, \ldots, \mu_K)$. The kth distribution generates sample $x_{k,t}$ at time t, for integer $t \geq 1$, and all of the samples $x_{k,t}$ are independent. A policy π is a function that, at each time t, chooses the index k_t of a distribution to sample based on the previously observed samples. That is,

$$k_t = \pi(\underbrace{k_1, \ldots, k_{t-1}}_{\text{past arm choices}}, \underbrace{x_{k_1,1}, \ldots, x_{k_{t-1},t-1}}_{\text{past values}}).$$

We would like to compare the performance of a policy π to that of an oracle policy π_* that has access to knowledge of the distributions μ_1^K, so

$$k_t^* = \pi_*(\mu_1^K, k_1^*, \ldots, k_{t-1}^*, x_{k_1^*,1}, \ldots, x_{k_{t-1}^*,t-1}).$$

Both Cicirello and Smith [6] and Carpentier and Valko [5] phrase their results in terms of the maximization of a reward rather than the minimization of a cost. They define the "regret" of policy π with respect to the oracle π_* over a time horizon of T as

$$G_T^{\pi,\pi_*} = \mathbb{E}\left[\max_{1 \leq t \leq T} x_{k_t^*,t}\right] - \mathbb{E}\left[\max_{1 \leq t \leq T} x_{k_t,t}\right].$$

Under semiparametric assumptions on μ_1^K, Carpentier and Valko [5] exhibit a policy π satisfying

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\max_{1\le t\le T} x_{k_t,t}\right]}{\mathbb{E}\left[\max_{1\le t\le T} x_{k_t^*,t}\right]} \to 1. \tag{1}$$

Equation 1 is equivalent to the more familiar statement that G_T^{π,π_*} is $o(\mathbb{E}\left[\max_{1\le t\le T} x_{k_t^*,t}\right])$, and it suggests that the policy π performs asymptotically as well as the oracle. While the condition in Equation 1 is sensible for the setting considered by Carpentier and Valko [5] (where the distributions μ_1^K have unbounded support), it is particularly sensitive to the nature of the distributions. For instance, the result in Equation 1 is trivially achieved when the distributions have bounded support (for example, in $[0,1]$ as in hyperparameter optimization). In this case, both the numerator and denominator converge to the upper bound of the support (for any policy that chooses each distribution infinitely often).

Furthermore, the condition in Equation 1 is assymetric with respect to maximization and minimization. When performing minimization of a cost instead of maximization of a reward (using distributions supported in $[0,1]$), both $\mathbb{E}\left[\min_{1\le t\le T} x_{k_t,t}\right]$ and $\mathbb{E}\left[\min_{1\le t\le T} x_{k_t^*,t}\right]$ may approach 0, in which case the ratio may exhibit radically different behavior. In Example 2 and Example 3, we demonstrate some of the peculiarities of this performance metric in the minimization setting.

Example 2. *Suppose μ_1 is a Bernoulli distribution with mean parameter $0 < p < 1$ and suppose that μ_2 is a point mass on 1. Consider a policy π which chooses μ_2 at $t = 1$ and then chooses μ_1 for all $t > 1$ and a policy π_* which always chooses μ_1. We have*

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\min_{1\le t\le T} x_{k_t,t}\right]}{\mathbb{E}\left[\min_{1\le t\le T} x_{k_t^*,t}\right]} = \lim_{T\to\infty} \frac{p^{T-1}}{p^T} = \frac{1}{p},$$

which remains bounded away from 1 even though the policy π acted optimally at every time step after $t = 1$.

Example 3. *Suppose μ_1 is the uniform distribution over $[0,1]$ and suppose that μ_2 is a point mass on 1. Consider a policy π which chooses μ_2 at $t = 1$ and then chooses μ_1 for all $t > 1$ and a policy π_* which always chooses μ_1. We have*

$$\lim_{T\to\infty} \frac{\mathbb{E}\left[\min_{1\le t\le T} x_{k_t,t}\right]}{\mathbb{E}\left[\min_{1\le t\le T} x_{k_t^*,t}\right]} = \lim_{T\to\infty} \frac{T^{-1}}{(T+1)^{-1}} \to 1.$$

Note above that the minimum of T independent uniform random variables is a Beta$(1,T)$ random variable, which has mean $1/(T+1)$.

Despite the fact that the policy π acts optimally at every time step other than $t = 1$ in both Example 2 and Example 3, the ratios of their expectations to that of the oracle π_* exhibit wildly different behaviors.

To avoid this sensitivity, we define regret as follows.

Definition 4. *We define the regret of the policy π with respect to the oracle policy π_* over a time horizon of T as*

$$R_T^{\pi,\pi_*} = \frac{1}{T} \min_{T'\ge 1} \left\{ T' : \mathbb{E}\left[\min_{1\le t\le T'} x_{k_t,t}\right] \le \mathbb{E}\left[\min_{1\le t\le T} x_{k_t^*,t}\right] \right\}.$$

Note that R_T^{π,π_*} depends on the tuple of distributions μ_1^K, but we suppress this dependence in our notation.

Then R_T^{π,π_*} is essentially the ratio of iterations that π requires to perform as well as the oracle π_* over a time horizon of T. This definition is sensible regardless of whether the samples are bounded or unbounded, whether we care about minimization or maximization, and regardless of how we scale or translate the distributions. Note that in both Example 2 and Example 3, we have $R_T^{\pi,\pi_*} = \frac{T+1}{T} \to 1$.

Definition 5. *We say that policy π achieves "no regret" with respect to the oracle π_* if $\limsup_T R_T^{\pi,\pi_*} \le 1$ for all tuples of distributions μ_1^K.*

Definition 5 is fairly lenient. Had we defined "no regret" using the condition given in Equation 1, our main result in Theorem 12 could have been made even stronger, but we view that as undesirable as illustrated by Example 2 and Example 3.

3 Oracle Models

In the standard multi-armed bandit setting, if an oracle with knowledge of the distributions of the arms seeks to minimize the expected sum of the losses, it should simply choose to play the arm with the lowest mean. This is true regardless of the time horizon. By anology with the usual multi-armed bandit setting, Cicirello and Smith [6] and Carpentier and Valko [5] both consider the oracle policy in Definition 6 that plays the single "best" arm.

Definition 6 (single-armed oracle)**.** *The single-armed oracle is the oracle, which over a time horizon of T, plays the single best arm*

$$\arg\min_k \mathbb{E}\left[\min_{1\le t\le T} x_{k,t}\right].$$

The single-armed oracle provides a good benchmark for comparison, but it is not the optimal oracle policy. When the time horizon is known in advance, the optimal oracle policy is given in Definition 7.

Definition 7 (optimal oracle)**.** *The optimal oracle over a time horizon of T plays the policy that solves*

$$\arg\min_\pi \mathbb{E}\left[\min_{1\le t\le T} x_{k_t,t}\right].$$

When the time horizon is not known in advance, one possible oracle strategy is a greedy strategy given in Definition 8.

Definition 8 (greedy oracle)**.** *The greedy oracle chooses the arm k_t at time t that gives the maximal expected improvement over the current best $y_{t-1} = \min_{1\le s\le t-1} x_{k_s,s}$. That is,*

$$k_t = \arg\min_k \mathbb{E}\left[\min\{x_{k,t}, y_{t-1}\} \mid x_{k_1,1}, \ldots, x_{k_{t-1},t-1}\right].$$

Unlike the greedy oracle, both the single-armed oracle and the optimal oracle require knowledge of the time horizon. Indeed, as shown in Example 9, the notion of a "best" arm is not well-defined outside of a specific time horizon. The best arm depends on the time horizon. This point contrasts sharply with the usual multi-armed bandit setting.

Example 9. *Suppose we have an infinite collection of arms μ_s indexed by $0 < s < 1$. Let $x_{s,t}$ be a sample from μ_s such that $P(x_{s,t} = s) = s$ and $P(x_{s,t} = 1) = 1 - s$. Then the optimal s is $O((\log T)/T)$.*

We elaborate on Example 9 in Appendix A. One difference between the single-armed oracle and the optimal oracle is that the optimal oracle can adapt its strategy based on the samples that it receives, whereas the single-armed oracle is non-adaptive. Its strategy is fixed ahead of time. Example 10 shows that the single-armed oracle is not even the optimal non-adaptive oracle. A mixed strategy may outperform any policy that plays only a single arm.

Example 10. *Consider a time horizon $T = 2$ and consider two arms. Suppose that samples $x_{1,t}$ from μ_1 deterministically equal $1/2$ and that samples $x_{2,t}$ from μ_2 satisfy $P(x_{2,t} = 0) = 1/4$ and $P(x_{2,t} = 1) = 3/4$. Then*

$$\mathbb{E}\min_{1\le t\le 2} x_{1,t} = \frac{1}{2} \qquad \mathbb{E}\min_{1\le t\le 2} x_{2,t} = \frac{9}{16} \qquad \mathbb{E}\min\{x_{1,1}, x_{2,2}\} = \frac{3}{8}.$$

This example shows that a fixed strategy that plays both arms can outperform any policy that plays a single-arm.

We described three different oracle models above. One caveat is that in the event that there is a well-defined best arm, that is, some arm k_* such that $P(x_{k_*,t} \le \alpha) \ge P(x_{k,t} \le \alpha)$ for all k and all $0 \le \alpha \le 1$, then these three oracles all coincide and we need not worry about which oracle to use for comparison. This is roughly the case in prior work. Cicirello and Smith [6] and Carpentier and Valko [5] make (semi)parametric assumptions on the distributions of the arms which essentially restrict the setting to have a well-defined best arm.

Despite the fact that the single-armed oracle is not the optimal oracle strategy, it is often a sufficiently strong baseline for measuring the performance of our policies. When we cannot even do as well as the single-armed oracle, as will be the case in Theorem 12, then we also cannot do as well as the optimal oracle. For the remainder of the paper, we will compare to the single-armed oracle. However, the results necessarily hold for comparisons to the optimal oracle as well.

4 Main Result

Theorem 12 shows that no policy can be guaranteed to perform asymptotically as well as the single best distribution. That is, it is impossible to achieve "no regret" in the extreme bandit problem. Moreover, in the worst case, no policy can asymptotically outperform the policy of choosing each distribution equally often, at least not without further assumptions.

Remark 11. *If π_{eq} is the policy that chooses $k_t \equiv t \bmod K$ (that is, it chooses each arm equally often) and π_* is the single-armed oracle from Definition 6, then* $\limsup_T R_T^{\pi_{eq},\pi_*} \leq K$ *regardless of the distributions $\mu_1^K = (\mu_1, \dots, \mu_K)$.*

We show in Theorem 12 that no policy π can improve on the guarantee of π_{eq} given in Remark 11.

Theorem 12. *For any policy π, there exist distributions μ_1^K such that* $\limsup_T R_T^{\pi,\pi_*} \geq K$*, where π_* is the single-armed oracle.*

We prove Theorem 12 in Section 4.3. The main components of the proof are Lemma 14, which upper bounds the performance of the single-armed oracle and Lemma 16, which lower bounds the performance of the policy π.

This result shows that the extreme bandit problem is fundamentally different from the standard multi-armed bandit problem, where a variety of policies perform asymptotically as well as the single best arm. Indeed, in the standard bandit problem, the arms are primarily characterized by their means, and so it suffices to estimate the means of the arms and play the best one. However, as discussed in Example 9, there is no well-defined best arm in the extreme bandit problem. Our construction will create a situation where the "best" arm periodically switches among the K distributions so that the policy π often ends up choosing the "wrong" arm.

For $i \geq 1$, let $\alpha_i = (8K)^{-(i!)^2}$. Our construction will involve a sum of point masses at the values α_i. It is easily verified that the sequence α_i satisfies the conditions in Lemma 13.

Lemma 13. *The sequence α_i satisfies the following properties.*

- *(A)* $\sum_{j=1}^{\infty} \alpha_j \leq 1/2$
- *(B)* $\alpha_i \leq \frac{1}{4(1+i)}$
- *(C)* $\sum_{j=i+1}^{\infty} \alpha_j \leq \frac{\alpha_i}{iK}$
- *(D)* $\alpha_i \leq \alpha_{i-1}^i 2^{-i}$.

Henceforth, we will not need the exact values of the sequence, we will only need the properties enumerated in Lemma 13. For $b = (b_1, b_2, \dots) \in \{1, \dots, K\}^\infty$, define the tuple of distributions $\mu_1^K(b) = (\mu_1(b), \dots, \mu_K(b))$ via

$$\mu_k(b) = \gamma_k(b)\delta_1 + \sum_{i=1}^{\infty} \mathbb{1}[b_i = k]\, \alpha_i\, \delta_{\alpha_i} \quad \text{where} \quad \gamma_k(b) = 1 - \sum_{i=1}^{\infty} \mathbb{1}[b_i = k]\alpha_i.$$

Here, δ_c represents a point mass at c, $\mathbb{1}[\xi]$ is the $\{0,1\}$-indicator function of the event ξ, and $\gamma_k(b)$ is chosen to make $\mu_k(b)$ a probability measure. Let M_K be the set of tuples of distributions that can be obtained in this way. The value b_i simply assigns the point mass

δ_{α_i} to one of the K distributions. We let D denote the distribution over the set $\{1, \ldots, K\}^\infty$ defined so that the b_i's are independent uniform random variables in $\{1, \ldots, K\}$.

Define the time horizon $T_i = \lceil \log(1/\alpha_i)/\alpha_i \rceil$. Instead of controlling $R_T^{\pi;\pi_*}$ for every T, we will control the quantity specifically for the time horizons T_i. In our construction, the b_ith arm in the tuple will be the best arm over the time horizon T_i, and the other arms will be substantially worse. We will show that, for a fixed i, we can construct a tuple μ_1^K so that the policy π takes roughly K times longer than the single-armed oracle π_* to obtain the value α_i (that is, π_* requires roughly T_i samples and π requires roughly $T_i' \approx KT_i$ samples). Using the probabilistic method, we will then show that we can find a tuple μ_1^K so that the policy takes roughly K times longer than the oracle to obtain the value α_i for infinitely many values of i.

4.1 Upper Bound on Oracle Performance

We begin by giving an upper bound on the performance of the oracle policy that plays the single best arm over the time horizon T_i. This bound holds uniformly over M_K.

Lemma 14. *Suppose that $\mu_1^K(b) \in M_K$. If π_* is the single-armed oracle from Definition 6, then*

$$\mathbb{E}\left[\min_{t \le T_i} x_{k_*,t}\right] < 2\alpha_i.$$

Proof. Recall that b_i is the index of the distribution that has a point mass at α_i. We have

$$\mathbb{E}\left[\min_{t \le T_i} x_{k_*,t}\right] = \min_k \mathbb{E}\left[\min_{t \le T_i} x_{k,t}\right] \le \mathbb{E}\left[\min_{t \le T_i} x_{b_i,t}\right].$$

The term on the right hand side can be rewritten as

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{1}\left[\min_{t \le T_i} x_{b_i,t} \le \alpha_i\right] \min_{t \le T_i} x_{b_i,t}\right] + \mathbb{E}\left[\mathbb{1}\left[\min_{t \le T_i} x_{b_i,t} > \alpha_i\right] \min_{t \le T_i} x_{b_i,t}\right] \\
&\le \alpha_i P\left[\min_{t \le T_i} x_{b_i,t} \le \alpha_i\right] + P\left[\min_{t \le T_i} x_{b_i,t} > \alpha_i\right] \\
&\le \alpha_i + P\left[\min_{t \le T_i} x_{b_i,t} > \alpha_i\right].
\end{aligned}$$

The first inequality follows by upperbounding the term $\min_{t \le T_i} x_{b_i,t}$ by α_i in the first term and by 1 in the second term. The second inequality follows by upperbounding the first probability by 1. To finish the lemma, note that

$$P\left[\min_{t \le T_i} x_{b_i,t} > \alpha_i\right] \le (1 - \alpha_i)^{T_i} < e^{-\alpha_i T_i} \le \alpha_i,$$

where the third inequality uses the definition $T_i = \lceil \log(1/\alpha_i)/\alpha_i \rceil$. □

4.2 Lower Bound on Performance of π

Here, we give a lower bound on the performance of any fixed policy π, when averaged over a collection of tuples of distributions.

Define the time horizon $T_i' = \lfloor c_i K \log(1/\alpha_i)/\alpha_i \rfloor$, where $c_i = (1 - 1/i)/((1 + 1/i)^2 + 2/i)$. The constant c_i is a correction term that converges to 1 as $i \to \infty$. Its specific value is not meaningful. The goal of this section is roughly to show that the performance of the policy π over a time horizon of T_i' is comparable to the performance of the oracle policy over a time horizon of T_i.

Throughout this section, we will fix an index i and we fix b_j for all $j \ne i$. Then we define the sequence $b^{k'} = (b_1^{k'}, b_2^{k'}, \ldots)$ via $b_j^{k'} = b_j$ for $j \ne i$ and $b_i^{k'} = k'$. The K tuples $\mu_1^K(b^{k'})$

for different values of k' are identical in all respects except for the index of the distribution that possesses the point mass δ_{α_i} and the amount of mass $\gamma_k(b^{k'})$ that the kth distribution in the k'th tuple assigns to δ_1.

Define the tuple of distributions $\eta_1^K(\overline{b}) = (\eta_1(\overline{b}), \ldots, \eta_K(\overline{b}))$ by $\eta_k(\overline{b}) = \frac{1}{K}\sum_{k'=1}^K \mu_k(b^{k'})$. Let $\gamma_k(\overline{b}) := \frac{1}{K}\sum_{k'=1}^K \gamma_k(b^{k'})$ denote the probability that $\eta_k(\overline{b})$ assigns to the value 1. The tuple $\eta_1^K(\overline{b})$ is the average of the tuples $\mu_1^K(b^{k'})$ over the different values of k'.

We begin with Lemma 15 which compares the probability that policy π obtains the value α_i when averaged over the tuples $\mu_1^K(b^{k'})$ with the probability that π obtains the value α_i in the tuple $\eta_1^K(\overline{b})$. This comparison is helpful because each distribution in the tuple $\eta_1^K(\overline{b})$ assigns the same mass of α_i/K to α_i and so the probability that π obtains α_i when run on the tuple $\eta_1^K(\overline{b})$ does not depend on π (it is simply $(1-\alpha_i/K)^{T_i'}$ where T_i' is the time horizon). Of course, as stated, we are actually concerned with the probability that π obtains a value less than or equal to α_i, but because of Lemma 13(C), the contribution of the smaller terms will not be too great.

Lemma 15. *We have*

$$\frac{1}{K}\sum_{k'=1}^K P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \mu_1^K(b^{k'})\right] \ge cP\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \eta_1^K(\overline{b})\right],$$

where $c = e^{-\frac{2\alpha_i T_i'}{iK}}$. In our notation, we condition on $\mu_1^K(b^{k'})$ to indicate the tuple of distributions being used.

Proof. Define $S(\pi, \mu_1^K, T)$ to be the set of actions and values that can be obtained by following policy π on the tuple μ_1^K for a time horizon of T. That is,

$$S(\pi, \mu_1^K, T) = \left\{(k_t, x_t)_{t=1}^T \,:\, k_t = \pi(k_1, \ldots, k_{t-1}, x_1, \ldots, x_{t-1}),\ x_t \in \mathrm{supp}(\mu_{k_t})\right\},$$

where $\mathrm{supp}(\mu_{k_t})$ is the support of the distribution μ_{k_t}. Then define $S(\pi, \mu_1^K, T, i)$ to be the subset of $S(\pi, \mu_1^K, T)$ such that all values are greater than or equal to α_{i-1}. That is,

$$S(\pi, \mu_1^K, T, i) = \left\{(k_t, x_t)_{t=1}^T \in S(\pi, \mu_1^K, T) \,:\, x_t \ge \alpha_{i-1}\right\}.$$

Critically, note that

$$S(\pi, \mu_1^K(b^1), T_i', i) = \cdots = S(\pi, \mu_1^K(b^K), T_i', i) = S(\pi, \eta_1^K(\overline{b}), T_i', i). \tag{2}$$

Equation 2 holds because the supports of the tuples $\mu_1^K(b^{k'})$ and $\eta_1^K(\overline{b})$ only differ on α_i, but we are considering only values that are at least α_{i-1}, so this difference does not affect the sets. We shall refer to this common set as S. We have

$$P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \mu_1^K(b^{k'})\right] = \sum_S \left(\prod_{j=1}^{i-1} \alpha_j^{|\{t\,:\,x_t=\alpha_j\}|} \prod_{k=1}^K \gamma_k(b^{k'})^{|\{t\,:\,k_t=k, x_t=1\}|}\right). \tag{3}$$

It follows that

$$\begin{aligned}
&\frac{1}{K}\sum_{k'=1}^K P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \mu_1^K(b^{k'})\right] \\
&= \frac{1}{K}\sum_{k'=1}^K \sum_S \left(\prod_{j=1}^{i-1} \alpha_j^{|\{t\,:\,x_t=\alpha_j\}|} \prod_{k=1}^K \gamma_k(b^{k'})^{|\{t\,:\,k_t=k, x_t=1\}|}\right) \qquad (4)\\
&= \sum_S \left(\prod_{j=1}^{i-1} \alpha_j^{|\{t\,:\,x_t=\alpha_j\}|} \left(\frac{1}{K}\sum_{k'=1}^K \prod_{k=1}^K \gamma_k(b^{k'})^{|\{t\,:\,k_t=k, x_t=1\}|}\right)\right),
\end{aligned}$$

where the first equality uses Equation 3 and the second equality simply rearranges the terms. We would like to essentially apply Jensen's inequality to say something like

$$\frac{1}{K}\sum_{k'=1}^K \prod_{k=1}^K \gamma_k(b^{k'})^{|\{t\,:\,k_t=k, x_t=1\}|} \ge \prod_{k=1}^K \gamma_k(\overline{b})^{|\{t\,:\,k_t=k, x_t=1\}|}. \tag{5}$$

Unfortunately, despite the fact that γ_k is convex on the relevant region, $\prod_{k=1}^K \gamma_k$ is not quite convex. However, it is nearly convex, and as we show in Lemma 18, Equation 5 holds up to a correction factor of $e^{-\frac{2\alpha_i T_i'}{iK}}$. Using this result in Equation 4 gives

$$\begin{aligned}
&\frac{1}{K}\sum_{k'=1}^{K} P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \mu_1^K(b^{k'})\right] \\
&\ge e^{-\frac{2\alpha_i T_i'}{iK}} \sum_S \left(\prod_{j=1}^{i-1} \alpha_j^{|\{t\,:\,x_t=\alpha_j\}|} \prod_{k=1}^{K} \gamma_k(\overline{b})^{|\{t\,:\,k_t=k, x_t=1\}|}\right) \\
&= e^{-\frac{2\alpha_i T_i'}{iK}} P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \eta_1^K(\overline{b})\right].
\end{aligned}$$

the first inequality uses Lemma 18 and the last equality holds for the same reason that Equation 3 holds. ☐

In Lemma 16, we turn the bound in Lemma 15 on the probability of obtaining α_i into a bound on the performance of π. Note that Lemma 16 holds uniformly over the values of b_j for $j \neq i$.

Lemma 16. *We have*

$$\frac{1}{K}\sum_{k'=1}^{K} \mathbb{E}\left[\min_{t\le T_i'} x_{k_t,t} \,\middle|\, \mu_1^K(b^{k'})\right] \ge 2\alpha_i.$$

Proof. We have

$$\begin{aligned}
\frac{1}{K}\sum_{k'=1}^{K} \mathbb{E}\left[\min_{t\le T_i'} x_{k_t,t} \,\middle|\, \mu_1^K(b^{k'})\right] &\ge \frac{\alpha_{i-1}}{K}\sum_{k'=1}^{K} P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \mu_1^K(b^{k'})\right] \\
&\ge \alpha_{i-1} e^{-\frac{2\alpha_i T_i'}{iK}} P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \eta_1^K(\overline{b})\right]
\end{aligned} \tag{6}$$

The first inequality is Markov's inequality. The second inequality is Lemma 15. We have

$$\begin{aligned}
P\left[\min_{t\le T_i'} x_{k_t,t} \ge \alpha_{i-1} \,\middle|\, \eta_1^K(\overline{b})\right] &\ge \left(1 - \frac{\alpha_i}{K} - \sum_{j=i+1}^{\infty} \alpha_j\right)^{T_i'} \\
&\ge \left(1 - \frac{\alpha_i(1+\frac{1}{i})}{K}\right)^{T_i'} \\
&\ge e^{-\alpha_i(1+\frac{1}{i})^2 T_i'/K} \\
&\ge \alpha_i^{(1+\frac{1}{i})^2 c_i}.
\end{aligned} \tag{7}$$

The first inequality lower bounds the probability of obtaining a value of α_i or less at every iteration. The second inequality uses Lemma 13(C). The third inequality uses Lemma 19 and Lemma 13(B). The fourth inequality uses the definition $T_i' = \lfloor c_i K \log(1/\alpha_i)/\alpha_i \rfloor$. Combining the Equation 6 and Equation 7 gives

$$\begin{aligned}
\frac{1}{K}\sum_{k'=1}^{K} \mathbb{E}\left[\min_{t\le T_i'} x_{k_t,t} \,\middle|\, \mu_1^K(b^{k'})\right] &\ge \alpha_{i-1} e^{-\frac{2\alpha_i T_i'}{iK}} \alpha_i^{(1+\frac{1}{i})^2 c_i} \\
&\ge 2\alpha_i^{\frac{1}{i}} \alpha_i^{\frac{2c_i}{i}} \alpha_i^{(1+\frac{1}{i})^2 c_i} \\
&= 2\alpha_i.
\end{aligned}$$

The second inequality uses Lemma 13(D) and the definition of T_i'. The third line uses the definition $c_i = (1 - 1/i)/((1 + 1/i)^2 + 2/i)$, which was chosen to make the third line hold. This completes the proof of the lemma. ☐

Noting that Lemma 16 holds uniformly over the values of b_j for $j \neq i$, a direct consequence of Lemma 16 is Corollary 17.

Corollary 17. *We have*

$$P_{b\sim D}\left(\mathbb{E}\left[\min_{t\leq T_i'} x_{k_t,t}\middle|\mu_1^K(b)\right] \geq 2\alpha_i\right) \geq \frac{1}{K},$$

where D is the distribution over $\{1,\ldots,K\}^\infty$ defined by sampling each component independently and uniformly at random from $\{1,\ldots,K\}$. The outer propability is over b, and the inner expectation is over the $x_{k_t,t}$.

4.3 Proof of Theorem 12

Here we synthesize the above results to prove Theorem 12. Lemma 14 and Corollary 17 together imply that

$$P_{b\sim D}\left(\mathbb{E}\left[\min_{t\leq T_i'} x_{k_t,t}\,\middle|\,\mu_1^K(b)\right] \geq 2\alpha_i > \mathbb{E}\left[\min_{t\leq T_i} x_{k_*,t}\,\middle|\,\mu_1^K(b)\right]\right) \geq \frac{1}{K},$$

which directly implies that $P(R_{T_i}^{\pi,\pi_*} \geq T_i'/T_i) \geq 1/K$. Recall that for a sequence of events A_i, we have $P(\text{infinitely many}\, A_i\, \text{happen}) \geq \limsup P(A_i)$. This can be seen by applying Fatou's lemma to the relevant indicator functions. It follows that

$$P_{b\sim D}\left(R_{T_i}^{\pi,\pi_*} \geq \frac{T_i'}{T_i}\,\text{for infinitely many}\, i\right) \geq \frac{1}{K}.$$

Recall the definitions

$$T_i = \lceil \log(1/\alpha_i)/\alpha_i \rceil \qquad\qquad T_i' = \lfloor c_i K \log(1/\alpha_i)/\alpha_i \rfloor.$$

Since $c_i \to 1$, it follows that $T_i'/T_i \to K$, and so there exists a tuple $\mu_1^K \in M_K$ such that $\limsup_T R_T^{\pi,\pi_*} \geq K$, proving the claim.

5 Related Work

Our setting is closely related to the multi-armed bandit problem, which has been studied extensively. See Bubeck and Cesa-Bianchi [3] for a survey. Regret is the most common measure of performance, though some authors study "simple regret" [4], where the goal is to identify the arm with the lowest mean. However, these settings provide little guidance on designing a policy to minimize the single smallest cost. The extreme bandit problem, where we care not about the average cost but about the single minimal cost, has been significantly less studied.

The extreme bandit problem (also called the max K-armed bandit problem) is introduced in Cicirello and Smith [6] and further developed in Streeter and Smith [12, 13]. The problem is additionally studied in Carpentier and Valko [5], where the authors give an explicit algorithm and prove that it exhibits asymptotically no regret in the sense of Equation 1. However, all results in previous work have relied heavily on strong parametric or semiparametric assumptions on the distributions μ_1^K under consideration. Motivated by extreme value theory, Cicirello and Smith [6] assume that the distributions belong to the Gumbel family and Carpentier and Valko [5] consider distributions in the Fréchet family (or distributions that are well approximated by the Fréchet family). When the individual samples arise as the maxima of a large number of independent, identically-distributed random variables, then these assumptions may be realistic. These assumptions dramatically simplify the problem. As in the multi-armed bandit setting, where every sample from a distribution provides information about the mean of the distribution, in the parametric setting, every sample provides information about the parameters of the distribution. Once we have accurately estimated each distribution, we can make sensible choices about which distribution to choose.

Our work shows that some form of assumptions are necessary to improve on the guarantees of the policy that chooses each arm equally often.

We do not expect the parametric assumptions motivated by extreme value theory to make sense in the setting of hyperparameter optimization. However, the question of what realistic assumptions are likely to hold in practice for hyperparameter optimization is an important question.

The no free lunch theorems are another form of hardness result in the optimization setting. Wolpert and Macready [15] show that in a discrete setting, all optimization algorithms that never revisit the same point perform equally well in expectation with respect to the uniform distribution over all possible objectives.

6 Discussion

We have shown that no policy can be guaranteed to perform better than the policy of choosing each distribution equally often. This result should not be construed to say that no policy can do better in practice. Indeed, hyperparameter optimization problems in the real world possess many nice structural properties. For instance, many hyperparameters have a sweet spot outside of which the algorithm performs poorly. This suggests that many black-box objectives for hyperparameter optimization may exhibit coordinate-wise quasiconvexity. Crafting plausible assumptions on the objectives and understanding how they translate into conditions on the induced distributions over algorithm performance is an important problem.

Acknowledgments

We would like to thank Balázs Kégl for valuable discussions.

References

[1] J. Bergstra and Y. Bengio. Random search for hyper-parameter optimization. *The Journal of Machine Learning Research*, 13(1):281–305, 2012.

[2] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, pages 2546–2554, 2011.

[3] S. Bubeck and N. Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.

[4] S. Bubeck, R. Munos, and G. Stoltz. Pure exploration in finitely-armed and continuous-armed bandits. *Theoretical Computer Science*, 412(19):1832–1852, 2011.

[5] A. Carpentier and M. Valko. Extreme bandits. In *Advances in Neural Information Processing Systems*, pages 1089–1097, 2014.

[6] V. A. Cicirello and S. F. Smith. The max K-armed bandit: A new model of exploration applied to search heuristic selection. In *Proceedings of the National Conference on Artificial Intelligence*, 2005.

[7] J. Duchi, M. Jordan, M. Wainwright, and A. Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015.

[8] N. Hansen. The CMA evolution strategy: a comparing review. In *Towards a New Evolutionary Computation. Advances on Estimation of Distribution Algorithms*, pages 75–102. Springer, 2006.

[9] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM Journal on Optimization*, 9(1):112–147, 1998.

[10] M. Powell. The NEWUOA software for unconstrained optimization without derivatives. In *Large-Scale Nonlinear Optimization*, volume 83, pages 255–297. Springer, 2006.

[11] J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*, pages 2951–2959, 2012.

[12] M. J. Streeter and S. F. Smith. An asymptotically optimal algorithm for the max K-armed bandit problem. In *Proceedings of the National Conference on Artificial Intelligence*, 2006.

[13] M. J. Streeter and S. F. Smith. A simple distribution-free approach to the max K-armed bandit problem. In *Principles and Practice of Constraint Programming-CP 2006*, pages 560–574, 2006.

[14] Z. Wang, M. Zoghi, F. Hutter, D. Matheson, and N. de Freitas. Bayesian optimization in high dimensions via random embeddings. In *International Joint Conferences on Artificial Intelligence*, 2013.

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

A The Best Arm Depends on the Time Horizon

In Example 9, we considered an infinite collection of arms μ_s indexed by $0 < s < 1$. Samples $x_{s,t}$ from μ_s satisfy $P(x_{s,t} = s) = s$ and $P(x_{s,t} = 1) = 1 - s$. We claimed that for a time horizon of T, the optimal s is $O((\log T)/T)$.

We have

$$\mathbb{E}\left[\min_{1\le t\le T} x_{s,t}\right] = s(1-(1-s)^T) + 1(1-s)^T = s + (1-s)^{T+1}.$$

Let s_* be the index of the optimal distribution, so $\min_s \mathbb{E}[\min_{1\le t\le T} x_{s_*,t}] = s_* + (1-s_*)^{T+1}$. For large T, we can consider the range $0 < s \le \frac{1}{2}$. We have

$$s + e^{-2s(T+1)} \le s + (1-s)^{T+1} \le s + e^{-s(T+1)}.$$

It follows that

$$s_* + e^{-2s_*(T+1)} \le \min_s \mathbb{E}\left[\min_{1\le t\le T} x_{s,t}\right] \le \min_s s + e^{-s(T+1)} \le \frac{\log T}{T+1} + \frac{1}{T} \le \frac{2\log T}{T}.$$

Therefore, $s_* \le (2\log T)/T$ and $e^{-2s_*(T+1)} \le (2\log T)/T$. The latter implies that

$$s_* \ge \frac{-\log 2 - \log\log T + \log T}{2(T+1)}$$

These results imply that s_* is $O((\log T)/T)$.

B Proof of Lemma 18

Here we state and prove Lemma 18, which is used in the proof of Lemma 15. The goal of Lemma 18 is to show that the probability of a particular sequence of values under the tuple $\mu_1^K(b^{k'})$, when averaged over the possible values of k', is at least as great (up to a constant c) as the probability of the same sequence of values under the averaged tuple η_1^K. Since all values other than the values 1 and α_i have equal probability under all tuples (for $j \neq i$, the value α_j has probability α_j under the b_jth element of each tuple), this lemma focuses on the probabilities of the values that equal 1. Recall that $\gamma_k(b^{k'})$ is the probability of obtaining a value of 1 from $\mu_k(b^{k'})$ and $\gamma_k(\overline{b})$ is the probability of obtaining a value of 1 from η_k.

Lemma 18. *For integers $n_1, \ldots, n_K \geq 0$ such that $n_k \leq T$, we have*

$$\frac{1}{K}\sum_{k'=1}^{K}\prod_{k=1}^{K}\gamma_k(b^{k'})^{n_k} \geq c\prod_{k=1}^{K}\gamma_k(\bar{b})^{n_k},$$

where $c = e^{-\frac{2\alpha_i T}{iK}}$.

Proof. This result nearly follows from Jensen's inequality. Indeed, if the function

$$f(c_1, \ldots, c_K) = \prod_{k=1}^{K}\left(1 - c_k\alpha_i - \sum_{\substack{j=1\\ j\neq i}}^{\infty}\mathbb{1}[j=k]\alpha_j\right)^{n_k}$$

were convex, then the result would follow from a single application of Jensen's inequality. That is, the result with $c = 1$ is precisely the statement

$$\frac{f(1,0,\ldots,0) + \cdots + f(0,\ldots,0,1)}{K} \geq f\left(\frac{1}{K}, \ldots, \frac{1}{K}\right).$$

Unfortunately, despite the fact that f is the product of convex functions (over the relevant domains), f itself is not convex. To circumvent this difficulty, we will approximate each term with the exponential of an affine function, so that the product of approximations remains convex (because the affine functions simply add). As our approximation is imperfect, we pick up a penalty in the form of the constant c. Let

$$\omega_k = 1 - \sum_{\substack{j=1\\ j\neq i}}^{\infty}\mathbb{1}[j=k]\alpha_j \qquad \beta_{i,k} = \frac{\alpha_i}{\omega_k},$$

First write

$$\begin{aligned}\frac{1}{K}\sum_{k'=1}^{K}\prod_{k=1}^{K}\gamma_k(b^{k'})^{n_k} &= \frac{1}{K}\sum_{k'=1}^{K}\prod_{k=1}^{K}(\omega_k - \mathbb{1}[k'=k]\alpha_i)^{n_k}\\ &= \frac{1}{K}\left(\prod_{k=1}^{K}\omega_k^{n_k}\right)\sum_{k'=1}^{K}(1-\beta_{i,k'})^{n_{k'}}.\end{aligned} \tag{8}$$

Note that by Lemma 13(A), we have $\omega_{k'} \geq \frac{1}{2}$ and so $\beta_{i,k'} \leq 2\alpha_i$. It follows from Lemma 19 and Lemma 13(B) that we can write

$$\begin{aligned}\frac{1}{K}\sum_{k'=1}^{K}(1-\beta_{i,k'})^{n_{k'}} &\geq \frac{1}{K}\sum_{k'=1}^{K}e^{-(1+1/i)\beta_{i,k'}n_{k'}}\\ &\geq e^{-(1+1/i)\frac{1}{K}\sum_{k'=1}^{K}\beta_{i,k'}n_{k'}}\\ &\geq e^{-\frac{2\alpha_i T}{iK}}e^{-\frac{1}{K}\sum_{k'=1}^{K}\beta_{i,k'}n_{k'}}\\ &\geq e^{-\frac{2\alpha_i T}{iK}}\prod_{k'=1}^{K}\left(1-\frac{\beta_{i,k'}}{K}\right)^{n_{k'}}.\end{aligned} \tag{9}$$

The second inequality is Jensen's inequality. The third inequality breaks the $1 + 1/i$ term into two terms and uses the bounds $\beta_{i,k'} \leq 2\alpha_i$ and $n_{k'} \leq T$. The fourth inequality uses the fact that $e^{-x} \geq 1 - x$. Combining Equation 8 and Equation 9 gives

$$\begin{aligned}\frac{1}{K}\sum_{k'=1}^{K}\prod_{k=1}^{K}\gamma_k(b^{k'})^{n_k} &\geq e^{-\frac{2\alpha_i T}{iK}}\left(\prod_{k=1}^{K}\omega_k^{n_k}\right)\prod_{k'=1}^{K}\left(1-\frac{\beta_{i,k'}}{K}\right)^{n_{k'}}\\ &= e^{-\frac{2\alpha_i T}{iK}}\prod_{k=1}^{K}\left(\omega_k - \frac{\alpha_i}{K}\right)^{n_k}\\ &= e^{-\frac{2\alpha_i T}{iK}}\prod_{k=1}^{K}\gamma_k(\bar{b})^{n_k},\end{aligned}$$

which finishes the proof. ∎

C Upper Bound on Exponential

Throughout this paper, we make use of the inequality $e^{-x} \geq 1 - x$. However, on a couple of occasions, we need to lower bound $1 - x$ by an exponential of the form e^{-rx} for some constant r. The bound that we use is given in Lemma 19.

Lemma 19. *For $i \geq 1$ and $y \in [0, \frac{1}{2(1+i)}]$, we have $e^{-y(1+\frac{1}{i})} \leq 1 - y$.*

Proof. More generally, the convexity of e^{-x} implies that for $0 \leq x \leq c$, we have

$$e^{-x} \leq 1 - \frac{1 - e^{-c}}{c} x.$$

The right hand side is the formula for the line interpolating between the points $(0, 1)$ and (c, e^{-c}) on the graph of e^{-x}. Choosing $c = \log(1 + \frac{1}{i})$, and noting that $0 \leq x \leq \frac{1}{1+i}$ implies that $0 \leq x \leq c$ because of the standard inequality $1 - \frac{1}{x} \leq \log x$, we see that $0 \leq x \leq \frac{1}{1+i}$ implies that

$$e^{-x} \leq 1 - \frac{1 - \frac{i}{1+i}}{\log(1 + \frac{1}{i})} x \leq 1 - \frac{\frac{1}{1+i}}{\frac{1}{i}} x = 1 - \frac{i}{1+i} x.$$

Setting $y = \frac{i}{1+i} x$ and using the fact that $\frac{1}{2(1+i)} \leq \frac{i}{(1+i)^2}$ gives the result. $\square$