The Annals of Statistics
2023, Vol. 51, No. 4, 1850–1876
https://doi.org/10.1214/23-AOS2315

RELAXING THE I.I.D. ASSUMPTION: ADAPTIVELY MINIMAX OPTIMAL REGRET VIA ROOT-ENTROPIC REGULARIZATION

BY BLAIR BILODEAU[1,a], JEFFREY NEGREA[2,c] AND DANIEL M. ROY[1,b]

[1]*Department of Statistical Sciences, University of Toronto,* [a]*blair.bilodeau@mail.utoronto.ca,* [b]*daniel.roy@utoronto.ca*
[2]*Department of Statistics and Actuarial Science, University of Waterloo,* [c]*jnegrea@uwaterloo.ca*

We consider prediction with expert advice when data are generated from distributions varying arbitrarily within an unknown constraint set. This *semi-adversarial* setting includes (at the extremes) the classical i.i.d. setting, when the unknown constraint set is restricted to be a singleton, and the unconstrained adversarial setting, when the constraint set is the set of all distributions. The Hedge algorithm—long known to be minimax (rate) optimal in the adversarial regime—was recently shown to be simultaneously minimax optimal for i.i.d. data. In this work, we propose to relax the i.i.d. assumption by seeking adaptivity at all levels of a natural ordering on constraint sets. We provide matching upper and lower bounds on the minimax regret at all levels, show that Hedge with deterministic learning rates is suboptimal outside of the extremes and prove that one can adaptively obtain minimax regret at all levels. We achieve this optimal adaptivity using the follow-the-regularized-leader (FTRL) framework, with a novel adaptive regularization scheme that implicitly scales as the square root of the entropy of the current predictive distribution, rather than the entropy of the initial predictive distribution. Finally, we provide novel technical tools to study the statistical performance of FTRL along the semi-adversarial spectrum.

1. Introduction. In this work, we are concerned with obtaining guarantees for methods that make decisions in light of data. Often, such guarantees are obtained via assumptions on the distribution of data. One important example of such an assumption is that data are independent and identically distributed (i.i.d.). While this type of assumption—on the joint dependence structure of data—may be pragmatic, and can motivate methods that seem to perform well in practice, it is *impossible* to be sure that apparent structure observed in past data will continue. This impossibility highlights the inherent limitations of such assumptions: any guarantees about performance may fail in practice if the assumed dependency does not hold, and statistical methods that are optimal under a specific family of dependence structures may be far from optimal under another. It is of practical interest to determine when the performance of statistical methods is robust to the dependence structures that they are designed for, and to quantify how performance guarantees degrade as assumptions on the dependence structure are relaxed. Contrary to guarantees holding only under a specific dependence modelling assumption, meaningful guarantees should ideally hold regardless of the true nature of the data.

In order to study guarantees when even our most basic assumptions on data may not hold, we work in the setting of sequential decision making. In particular, we study the problem of *prediction with expert advice* [39, 57], where the statistician (decision maker) faces a sequence of prediction problems, on each round being offered the predictions of a finite set of "experts." On any round, the statistician's *regret* is defined to be the excess cumulative loss of

Received July 2022; revised July 2023.
MSC2020 subject classifications. Primary 62L10, 62C20; secondary 60G25, 62M20, 68Q32, 68T05.
Key words and phrases. Sequential decision theory, prediction with expert advice, adaptive minimax regret, aggregation, robust prediction.

the statistician's predictions, relative to the cumulative loss of the single best performing expert. Rather than minimize expected loss (a notion that may not be meaningful in this setting), the goal of the statistician is to minimize *regret*. In doing so, the statistician's performance is guaranteed to be, in aggregate, never much worse (and possibly better) than that of the single best expert in hindsight.

Prediction with expert advice is a classical problem, studied in statistics, information theory, theoretical computer science, game theory, finance and many other fields. A wide set of researchers has contributed to its development, beginning with the seminal work of Hannan [27], which introduces the concept of regret for sequential two player games, and of Cover [20], which studies the special case of two experts. Cesa-Bianchi and Lugosi [15] provide a detailed history of this problem in their book. In statistics, prediction with expert advice has close connections to empirical process theory [14] and statistical aggregation [3, 47, 53, 54]. Past work has studied prediction with expert advice under a wide range of assumptions on the data-generating mechanism, including when the prediction tasks (and expert advice) are assumed to be i.i.d. across rounds *as well as* when they are allowed to depend in an arbitrary way on past decisions.

Remarkably, there are meaningful minimax rates (of regret) for both the i.i.d. setting and the latter, so-called "adversarial," setting. By adopting a sequential prediction algorithm that is minimax in the adversarial setting, can we in a sense avoid making assumptions on the joint dependence structure of the sequence of prediction problems? On the one hand, adversarially minimax strategies come with nontrivial regret guarantees that hold without assumptions. On the other hand, these strategies can be very conservative and suffer much more regret than necessary when the data is (typical of) an i.i.d. realization. Somehow, we would like to provably exploit i.i.d. structure when it is present, without suffering too much regret when it is not.

One way to frame such guarantees is through the lens of *adaptation theory*, which has had a huge impact on statistics (see, e.g., [13]). In the sequential decision-making literature, so-called "best of both worlds" results demonstrate the existence of sequential prediction algorithms that simultaneously achieve minimax rates under both i.i.d. and adversarial settings. These results are not entirely satisfying, however, because they say nothing of performance when data are far from adversarial but not i.i.d. For example, at each round, our data may come from one of several distinct distributions, but we may not be confident to model the way that the distribution at each round depends on earlier rounds. If the populations are very similar, we would like our methods to perform almost as well as if the data were i.i.d.

In this work, we look closely at the performance of the HEDGE family of algorithms for prediction with expert advice, and study how performance degrades as we move away from i.i.d. environments. HEDGE algorithms maintain a generalized Bayesian posterior over the set of experts. Recent results [44] show that the DECREASING HEDGE variant can adapt in the i.i.d. setting to the ("stochastic") gap Δ between the expected loss of the "effective" (i.e., best) expert and other experts. We show that this and a broad class of variants of HEDGE *cannot* adapt to the presence of more than one effective expert, a natural generalization. (We call this the "semi-adversarial spectrum.") After identifying minimax rates in this new setting, we provide a novel algorithm, based on a new type of regularization, that adaptively achieves minimax rates at all deviations from i.i.d. along the semi-adversarial spectrum.

Our work contributes to a growing body of work that studies benign data-generating mechanisms without relying on the i.i.d. assumption (for a detailed survey, see Section 10.2). In particular, there is a large literature focused on developing decision procedures that satisfy data-dependent (random) regret bounds, and thus, adapt to quantities that could be seen to capture specific notions of regularity or easiness of the data. Examples of these quantities include the ℓ_∞ norm or empirical variance of the incurred losses. If one has a prior belief that these quantities will be small (e.g., the losses are expected to be highly correlated from round to round, likely leading to low empirical variance), one might opt to follow one of these decision procedures. The interpretation of data-dependent regret bounds is not always straightforward. Bounds based on the norm or variance of the losses do not necessarily offer strong guarantees for i.i.d. data with stochastic gaps: bounds that adapt to the empirical variance of the losses can be large when the observed losses vary significantly, yet this may occur even when the data is i.i.d., a setting for which smaller error bounds are possible.

In this work, we aim to adapt to data-generating mechanisms along a spectrum between i.i.d. and adversarial settings. In order to achieve this feat, we develop a novel data-dependent regret bound. By studying its expected value under various statistical assumptions, we demonstrate that our method achieves the desired minimax rates. In contrast with the alternative (and equally valid) adaptive regret bounds discussed above, i.i.d. settings (with a single best expert) are the "easiest" cases, and performance degrades smoothly as we relax the i.i.d. assumption toward the adversarial worst case. The semi-adversarial spectrum is one of many incomparable ways to quantify the relative easiness of decision making between i.i.d. and adversarial settings. While one should adopt a notion of easiness one expects to be relevant in one's domain, the semi-adversarial spectrum will be seen to have several desirable properties and applications.

Contributions and outline. In Section 2, we formally introduce prediction with expert advice in terms of a general, possibly adversarial, data-generating mechanism. Then, in Section 3, we define our relaxation of the i.i.d. assumption as an adversary constrained to a convex set of distributions, and introduce the quantities to which we will aim to adapt (the semi-adversarial spectrum). In Section 4, we formalize the relevant notion of adaptive minimax optimality and then summarize our main result: the existence of an adaptively minimax optimal prediction algorithm. In order to obtain upper bounds on regret, we derive a novel concentration-of-measure inequality for the semi-adversarial spectrum. This concentration inequality is presented in Section 5 and is likely of independent interest. In order to establish adaptive minimax optimality, we first identify the target minimax rates along the semi-adversarial spectrum. We do so, in Section 6, by deriving algorithm-agnostic lower bounds that match algorithm-specific upper bounds presented later for all scenarios along the semi-adversarial spectrum. In Section 7, we show DECREASING HEDGE and its variants are nonadaptive by constructing algorithm-specific lower bounds. The nonadaptive nature of these algorithms is connected to their nonadaptive learning rate tuning, as among D.HEDGE variants with nonadaptive tunings, only oracle-based tunings can yield minimax rates. In Sections 8 and 9, we introduce a novel algorithm, META-CARE, which implicitly and adaptively adjusts the learning rate of HEDGE without the need for oracle-based tuning, and prove that it is adaptively minimax optimal along the entire semi-adversarial spectrum. META-CARE operates by *boosting* our novel instance of the *follow-the-regularized-leader* (FTRL) algorithm, dubbed FTRL-CARE, with D.HEDGE using a second application of D.HEDGE, and hence a major component of our analysis is devoted to a general study of FTRL algorithms along the semi-adversarial spectrum, which comprises Section 8. We introduce and establish the adaptivity of META-CARE in Section 9. Finally, in Section 10, we conclude with a detailed review of related work. Technical details for the proofs of our results, and a brief simulation study, are available in the Appendices, which are available as the Supplementary Material [11].

2. Notation and problem setup.

Prediction with expert advice is characterized by the manner in which experts and the player make their predictions and the mechanism by which a response observation is generated. At every time $t \in \mathbb{N}$, each of the $N \in \mathbb{N}$ experts (arbitrarily indexed by $[N] = \{1, \dots, N\}$) formulate their predictions for the tth round, jointly denoted by $x(t) \in \hat{\mathcal{Y}}^N$, the player makes a prediction for the tth round, $\hat{y}(t) \in \hat{\mathcal{Y}}$, and the environment generates a response observation for the tth round, $y(t) \in \mathcal{Y}$. The *history* of the game up to time t is summarized by $h(t) = (x(s), \hat{y}(s), y(s))_{s\in[t]} \in \mathcal{H}^t$, where $\mathcal{H} = \hat{\mathcal{Y}}^N \times \hat{\mathcal{Y}} \times \mathcal{Y}$, with the convention that $h(0)$ is the empty tuple. For each time $t \in \mathbb{N}$, the prediction $\hat{y}(t)$ and response observation $y(t)$ are conditionally independent given the history $h(t-1)$ and the recent expert predictions, $x(t)$. This conditional independence reflects the fact that the player does not have access to the response until after making their prediction, and that the player has some private source of stochasticity with which to randomize their predictions.

The conditional distribution of the experts' predictions and the data observed at round t given $h(t-1)$ is uniquely described by a probability kernel $\pi_t \in \mathcal{K}(\mathcal{H}^{t-1}, \hat{\mathcal{Y}}^N \times \mathcal{Y})$, where $\mathcal{K}(\mathcal{A}, \mathcal{B})$ denotes the set of probability kernels (e.g., regular conditional distributions) from $\mathcal{A}$ to $\mathcal{B}$. Intuitively, π_t is the function describing how the data distribution on round t depends on the realized history up to that point. Letting $\mathscr{P}_N = \prod_{t\in\mathbb{N}} \mathcal{K}(\mathcal{H}^{t-1}, \hat{\mathcal{Y}}^N \times \mathcal{Y})$, a *data-generating mechanism* is any sequence $\pi = (\pi_t)_{t\in\mathbb{N}} \in \mathscr{P}_N$, which describes how the data distributions change over time in response to the observed history. Similarly, the conditional distribution of the player's prediction at time t given $h(t-1)$ and $x(t)$ is uniquely described by a probability kernel $\hat{\pi}_t \in \mathcal{K}(\mathcal{H}^{t-1} \times \hat{\mathcal{Y}}^N, \hat{\mathcal{Y}})$. Letting $\hat{\mathscr{P}}_N = \prod_{t\in\mathbb{N}} \mathcal{K}(\mathcal{H}^{t-1} \times \hat{\mathcal{Y}}^N, \hat{\mathcal{Y}})$, a *prediction policy* is any sequence $\hat{\pi} = (\hat{\pi}_t)_{t\in\mathbb{N}} \in \hat{\mathscr{P}}_N$, which describes how the player will behave over time in response to the observed history. Finally, since the $\hat{\pi}$ may be different depending on the number of experts (for example, the player may use a different learning rate), we provide a further level of generality describing how the player selects $\hat{\pi}$ as a function of the number of experts. More formally, a *prediction algorithm* is any sequence $\mathfrak{a} = (\hat{\pi}(N))_{N\in\mathbb{N}}$ with $\hat{\pi}(N) \in \hat{\mathscr{P}}_N$ for each N.

In a sequential prediction task, prior to any data being generated or predictions being made, the player selects a prediction algorithm and the environment determines a data-generating mechanism. Without loss of generality, the player knows the number of experts N, and so they predict according to the prediction policy $\hat{\pi} = \mathfrak{a}(N)$ based on their prediction algorithm. Due to the conditional independence assumption for $\hat{y}(t)$ and $y(t)$ given $h(t-1)$ and $x(t)$, the joint distribution of $(x(t), \hat{y}(t), y(t))_{t\in\mathbb{N}}$ is fully determined by the data-generating mechanism and the prediction policy selected by each party. For a data-generating mechanism π and a prediction policy $\hat{\pi}$, expectation under this joint law is denoted by $\mathbb{E}_{\pi,\hat{\pi}}$. When the prediction policy is determined by the prediction algorithm $\mathfrak{a}$, for any number of experts N and data-generating mechanism $\pi \in \mathscr{P}_N$ we use $\mathbb{E}_{\pi,\mathfrak{a}}$ to denote $\mathbb{E}_{\pi,\hat{\pi}}$, where $\hat{\pi} = \mathfrak{a}(N)$.

The accuracy of the player and experts is measured on each round using a loss function $\ell : \hat{\mathcal{Y}} \times \mathcal{Y} \to [0, 1]$, and the player's performance at the end of $T \in \mathbb{N}$ rounds of the game is measured by *regret*, defined as the $\sigma(h(T))$-measurable random variable

$$R(T) = \sum_{t=1}^{T} \ell\big(\hat{y}(t), y(t)\big) - \min_{i\in[N]} \sum_{t=1}^{T} \ell\big(x_i(t), y(t)\big).$$

In this work, we focus on bounding the expected regret $\mathbb{E}_{\pi,\hat{\pi}} R(T)$ for three specific prediction algorithms, so we use $\mathbb{E}_{\pi,\mathrm{H}}$, $\mathbb{E}_{\pi,\mathrm{C}}$ and $\mathbb{E}_{\pi,\mathrm{M}}$ to denote $\mathbb{E}_{\pi,\mathfrak{a}}$ under the D.HEDGE, FTRL-CARE and META-CARE algorithms, respectively (see Sections 7 to 9 for the respective definitions of these prediction algorithms).

Since $R(T)$ only depends on $h(T)$ through the loss function, expected regret bounds are often characterized using quantities that push the data-generating distributions forward through the loss function. Specifically, we define the *losses* $\ell_i(t) = \ell(x_i(t), y(t))$ and *cumulative losses* $L_i(t) = \sum_{s=1}^{t} \ell_i(s)$ for each expert $i \in [N]$ and $t \in \mathbb{N}$. Similarly, we define the *loss vector* $\ell(t) = (\ell_i(t))_{i\in[N]}$ and *cumulative loss vector* $L(t) = \sum_{s=1}^{t} \ell(s)$.

Let $\mathcal{M}(\mathcal{A})$ denote the set of all probability distributions on $\mathcal{A}$. For a distribution $\mu \in \mathcal{M}(\mathcal{A})$ and measurable function $f : \mathcal{A} \to \mathbb{R}$, we define $\mu f = \int_{\mathcal{A}} f(a)\mu(\mathrm{d}a)$. We will frequently use this notation for measures in $\mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$. In particular, for each expert $i \in [N]$, the expert's loss $\ell_i : (x, y) \mapsto \ell(x_i, y)$ is a function on $\hat{\mathcal{Y}}^N \times \mathcal{Y}$, and $\mu\ell_i$ is the expectation of expert i's loss when the expert predictions and response observation are jointly distributed as μ.

3. Semi-adversarial spectrum. We now introduce the *semi-adversarial* setting, in which we model how (non)adversarial a data-generating mechanism is via distributional constraints. At each round, conditional on the history, the expert's predictions $x(t)$ and the response observation $y(t)$ are sampled from a distribution lying in some set $\mathcal{D}$. When $\mathcal{D}$ is the entire probability simplex, the experts are unconstrained. Informally, as $\mathcal{D}$ "shrinks," the experts are more constrained and so competing with them becomes easier, from the player's perspective.

Formally, consider a fixed number of experts N. A *time-homogeneous convex constraint* is a convex subset $\mathcal{D} \subseteq \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$. Let $\mathscr{P}(\mathcal{D})$ denote the collection of data-generating mechanisms $\pi = (\pi_t)_{t \in \mathbb{N}}$ such that for all $t \in \mathbb{N}$ and $h \in \mathcal{H}^{t-1}$, $\pi_t(h, \cdot) \in \mathcal{D}$. That is, $\mathscr{P}(\mathcal{D})$ is the set of data-generating mechanisms under which the conditional distribution of the expert predictions and response data given the history is constrained to $\mathcal{D}$ but otherwise can vary arbitrarily within $\mathcal{D}$ depending on the history.

In Figure 1, we visualize possible trajectories of data-generating mechanisms for an i.i.d. environment, an adversarial environment and an environment whose constraint set is intermediate between these two extremes. In the i.i.d. case, the conditional distribution of the next instance given the history is fixed, and hence the constraint set corresponds to a single distribution on instances. In the adversarial case, the conditional distribution of the next instance given the history can vary arbitrarily in the space of all probability distributions on instances; in particular, it can be a point-mass at an adversarial instance for the player's strategy, depicted here as the extreme points of the space of distributions. Since the i.i.d. case corresponds to a singleton set of distributions on instances, and the adversarial case corresponds to the whole space of distributions on instances, a natural concept of "in between" these extremes is a proper subset of the set of distributions on instances. Our relaxation captures this by allowing the conditional distribution of the next instance given the history to vary within some convex constraint set that is not known by the player in advance (visualized here as an ellipse), and measuring performance relative to the properties of that unknown constraint set.

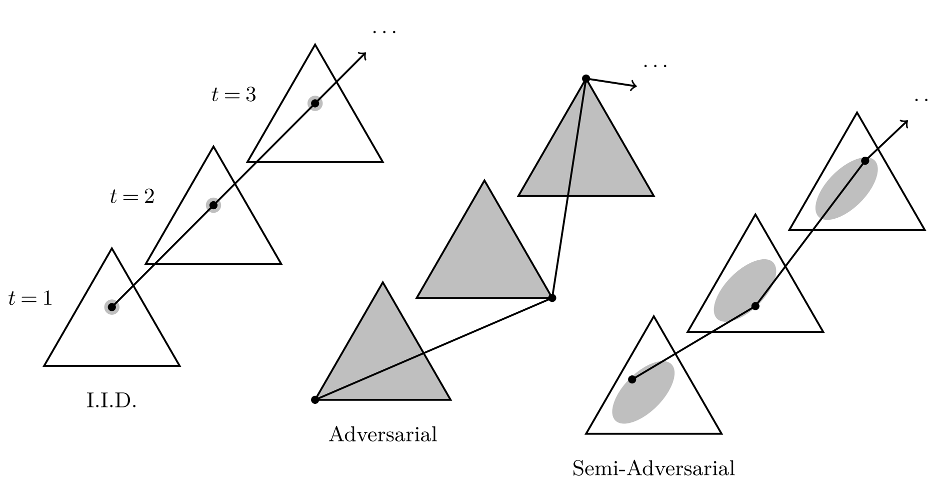

FIG. 1. *Visualizing the difference between i.i.d. data, adversarial data and a constraint set in between these two extremes. In each part of the figure, the triangles depict the set of conditional distributions for the tuple of expert predictions and response (an "instance") given the history at each time. The grey regions depict the space of conditional distributions for the next instance given the history that are possible for a given constraint set.*

In the context of prediction with expert advice, it is natural to assume that the time-homogeneous convex constraint $\mathcal{D}$ is unknown to the player. To what extent can we adapt to $\mathcal{D}$? In this work, we consider the problem of adapting to two characterizing quantities of $\mathcal{D}$: the *effective stochastic gap* and the *number of effective experts*. Formally, for each expert $i \in [N]$, let

$$\Delta_i(\mathcal{D}) = \inf_{\mu \in \mathcal{D}} \max_{i' \in [N]} \mu[\ell_i - \ell_{i'}],$$

and define the *effective stochastic gap* to be

$$\Delta_0(\mathcal{D}) = \min\{\Delta_i(\mathcal{D}) \mid i \in [N], \Delta_i(\mathcal{D}) > 0\},$$

with the convention $\min \varnothing = \infty$. Second, define the set of *effective experts*

$$\mathcal{I}_0(\mathcal{D}) = \{i \in [N] \mid \Delta_i(\mathcal{D}) = 0\},$$

and let $N_0(\mathcal{D}) = |\mathcal{I}_0(\mathcal{D})|$ denote the number of effective experts. The set $\mathcal{I}_0(\mathcal{D})$ contains the experts that *could be* the best (in conditional expectation given the history) on any particular round. Note that this set is always nonempty, since $\mathcal{D}$ contains at least one distribution and there is at least one expert who is optimal under this distribution.

The effective stochastic gap, $\Delta_0(\mathcal{D})$, is the minimal excess expected loss of an *ineffective expert* over the best effective expert on any round. If $\Delta_i(\mathcal{D}) > 0$, then the mean loss of expert i is always *strictly* larger than the mean loss of the best expert. Thus, $\Delta_0(\mathcal{D})$ corresponds to the smallest that this difference can be over all such experts and distributions. When $\mathcal{D}$ is clear, we simplify notation to $\mathcal{I}_0$, N_0 and Δ_0.

For a fixed N, N_0 and Δ_0, the collection of convex constraint sets that have these characterizing quantities is

$$\mathcal{V}(N, N_0, \Delta_0) = \{\mathcal{D} \subseteq \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y}) \mid \mathcal{D} \text{ convex}, N_0(\mathcal{D}) = N_0, \Delta_0(\mathcal{D}) \geq \Delta_0\},$$

and the corresponding set of data-generating mechanisms is

$$\mathscr{P}_{N,(N_0,\Delta_0)} = \bigcup_{\mathcal{D} \in \mathcal{V}(N,N_0,\Delta_0)} \mathscr{P}(\mathcal{D}).$$

Let $\mathscr{P} = \{\mathscr{P}_{N,(N_0,\Delta_0)} \mid N_0 \leq N \in \mathbb{N}, \Delta_0 > 0\}$ denote the collection of all such sets. Together, N_0 and Δ_0 induce a ranking of constraint sets, and the *semi-adversarial spectrum* is the collection of equivalence classes this ranking induces.

3.1. *Motivation for characterizing quantities.* Mourtada and Gaïffas [44] formalize the "best-of-both-worlds" adaptivity of D.HEDGE as simultaneously satisfying the adversarial minimax rate $\Theta(\sqrt{T \log N})$ and the minimax rates of the so-called "stochastic-with-a-gap" environments. Together, these environments are characterized by a known number of experts, N, and a single, unknown *stochastic gap*, denoted Δ. The stochastic-with-a-gap settings can be exactly captured by the characterizing quantities introduced above for the semi-adversarial spectrum: *The adversarial setting* corresponds to $\mathcal{D} = \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$ being the set of all distributions. Then $N_0(\mathcal{D}) = N$ and $\Delta_0(\mathcal{D}) = \infty$, and the minimax rate is $\Theta(\sqrt{T \log N})$. *In contrast, consider an i.i.d. setting*, where $\mathcal{D} = \{\mu_0\}$ is a singleton, for some $\mu_0 \in \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$; that is, for all $t \in \mathbb{N}$ and $h \in \mathcal{H}^{t-1}$, the data-generating mechanism satisfies $\pi_t(h, \cdot) = \mu_0$. It is straightforward to show that $\mathcal{I}_0(\{\mu_0\}) = \arg\min_{i \in [N]} \mu_0 \ell_i$ is the set of experts that are optimal (w.r.t. ℓ) in expectation under μ_0, each expert $i \in [N]$ has stochastic gap $\Delta_i(\{\mu_0\}) = \mu_0 \ell_i - \min_{i_0 \in [N]} \mu_0 \ell_{i_0}$ and the effective stochastic gap is $\Delta_0(\{\mu_0\}) = \min_{i \in [N] \setminus \mathcal{I}_0(\mu_0)} \Delta_i(\mu_0)$. The *stochastic-with-a-gap* setting, with stochastic gap $\Delta > 0$, is the set of all i.i.d. environments $\mathcal{D} = \{\mu_0\}$ with positive gap $\Delta_0(\{\mu_0\}) = \Delta > 0$ and one effective expert, that is, $N_0 = |\mathcal{I}_0(\mu_0)| = 1$. The minimax optimal expected regret (cf. [44]) is

$$\mathbb{E}R(T) \in \Theta\left(\frac{\log N}{\Delta}\right).$$

The semi-adversarial spectrum thus generalizes the stochastic-with-a-gap setting beyond one effective expert and generalizes the notion of stochastic gap to the notion of an effective stochastic gap. We will see that the dependence on N_0 and Δ_0 in the minimax rates along the semi-adversarial spectrum closely resembles the dependence on N and Δ in minimax rates in the adversarial and stochastic-with-a-gap settings.

In this work, we do not consider the possibility of adapting to a more refined spectrum than that induced by N_0 and Δ_0. As we will see, D.HEDGE cannot even adapt to this spectrum, and so there is no hope of it adapting to a further refinement. We will also see that the problem of adapting to N_0 and Δ_0 already leads us to a novel regularization scheme. We leave the study of alternative characterizing quantities and the more general problem of adapting to the finest possible spectrum to future work.

3.2. *Practical relevance of characterizing quantities.* A standard application of prediction with expert advice is to the setting of statistical aggregation (cf. [3, 32, 61]). Suppose the statistician has N models that map from a covariate space $\mathcal{X}$ to a response space $\mathcal{Y}$. Further, suppose that the t'th observation (X_t, Y_t) is sampled from one of K unknown "source" distributions on $\mathcal{X} \times \mathcal{Y}$, where K itself is also unknown. Rather than attempt to model how the choice of the source distribution for each observation depends on past observations, predictions, etc., we can treat the choices as adversarial (e.g., by using HEDGE), yielding guarantees without any such modeling assumptions. However, the (possibly randomized) selection of the source distribution by the data-generating mechanism gives rise to a time-homogeneous convex constraint $\mathcal{D}$, corresponding to the convex hull of the K unknown source distributions. If the source distributions and models are reasonably distinct, this will likely result in $N_0(\mathcal{D}) \leq K$ effective experts, which may be much smaller than N. META-CARE would automatically adapt in this situation, and scale with N_0, not N. Of course, were the statistician to, instead, impose misspecified modeling assumptions (e.g., derive predictions assuming the observations were i.i.d., when they were not), they might suffer linear regret.

More generally, for arbitrary (possibly nonconvex) constraint sets, taking the convex hull corresponds to allowing the data-generating mechanism access to biased coins with which to select between "basic" elements of the constraint set. Further, the regret against a nonconvex constraint set is no larger than the regret against the the convex hull of that constraint set, and hence our upper bounds still apply to nonconvex constraint sets. However, it is possible that the true lower bound for the nonconvex constraint set may no longer match our upper bounds. We leave a tight analysis of nonconvex lower bounds to future work.

3.3. *Examples of convex constraints.* The following examples illustrate the flexibility of time-homogeneous convex constraints and the semi-adversarial spectrum.

EXAMPLE 1 (I.I.D.-μ_0, Stochastic-with-a-gap). When the constraint set is the singleton $\mathcal{D}_{\mu_0} = \{\mu_0\}$, then there is only one possible data-generating mechanism, and under that data-generating mechanism the data and expert predictions are i.i.d. according to μ_0. Furthermore, if there exists $i_0 \in [N]$ and $\Delta > 0$ such that

$$\inf_{i \in [N] \setminus \{i_0\}} \mu[\ell_i - \ell_{i_0}] = \Delta$$

(i.e., there is a best expert in expectation under μ_0 and there is a gap of Δ from the best to the second best expert in expectation) then $\mathcal{I}_0(\mathcal{D}_{\mu_0}) = \{i_0\}$, $N_0(\mathcal{D}_{\mu_0}) = 1$ and $\Delta_0(\mathcal{D}_{\mu_0}) = \Delta$. This is called the *stochastic-with-a-gap* setting. Since any singleton is convex, $\mathcal{D}_{\mu_0}$ is convex.

EXAMPLE 2 (Adversarial). When the constraint set is the space of all probability measures $\mathcal{D}_{\text{adv}} = \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$, then the constrained setting reduces to the fully adversarial setting, since $\mathcal{D}$ contains all point-mass distributions. In this case, $\mathcal{I}_0(\mathcal{D}_{\text{adv}}) = [N]$, $N_0(\mathcal{D}_{\text{adv}}) = N$ and $\Delta_0(\mathcal{D}_{\text{adv}}) = +\infty$ (by convention, as it is the inf over an empty set). Since the set of all probability measures is convex, $\mathcal{D}_{\text{adv}}$ is convex.

EXAMPLE 3 (Adversarial-with-an-instantaneous-gap). For any $i_0 \in [N]$ and $\Delta \geq 0$,

$$\mathcal{D}_{i_0,\Delta}^{(\text{a.s.})} = \left\{\mu \in \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y}) \mid \mu\left(\ell_{i_0} + \Delta \leq \min_{i \in [N]\setminus\{i_0\}} \ell_i\right) = 1\right\}$$

is convex (since min is concave), and satisfies $\mathcal{I}_0(\mathcal{D}_{i_0,\Delta}^{(\text{a.s.})}) = \{i_0\}$, $N_0(\mathcal{D}_{i_0,\Delta}^{(\text{a.s.})}) = 1$ and $\Delta_0(\mathcal{D}_{i_0,\Delta}^{(\text{a.s.})}) = \Delta$. This contains all mixtures of point-mass distributions with common best expert i_0 that satisfy the gap constraint almost surely.

EXAMPLE 4 (Adversarial-with-an-$\mathbb{E}$-gap, Mourtada and Gaïffas [44]). For any $i_0 \in [N]$ and $\Delta \geq 0$,

$$\mathcal{D}_{i_0,\Delta} = \left\{\mu \in \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y}) \mid \mu\ell_{i_0} + \Delta \leq \min_{i \in [N]\setminus\{i_0\}} \mu\ell_i\right\}$$

is convex (since min is concave), and satisfies $\mathcal{I}_0(\mathcal{D}_{i_0,\Delta}) = \{i_0\}$, $N_0(\mathcal{D}_{i_0,\Delta}) = 1$ and $\Delta_0(\mathcal{D}_{i_0,\Delta}) = \Delta$. This relaxes the adversarial-with-an-instantaneous-gap setting, since $\mathcal{D}_{i_0,\Delta}^{(\text{a.s.})} \subseteq \mathcal{D}_{i_0,\Delta}$. This constraint set is equivalent to the formulation used in Corollary 6 of Mourtada and Gaïffas [44]; it is also the same setting as Section 4.2 of Wei and Luo [59], although they consider bandit feedback.

EXAMPLE 5 (Ball-around-I.I.D.). For any pseudometric d, radius $r > 0$ and probability measure μ_0,

$$\mathcal{D}_{\mu_0,d,r} = B_d(\mu_0, r) = \{\mu \in \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y}) \mid d(\mu, \mu_0) \leq r\}$$

is convex. The exact values of $\mathcal{I}_0(\mathcal{D}_{\mu_0,d,r})$, $N_0(\mathcal{D}_{\mu_0,d,r})$ and $\Delta_0(\mathcal{D}_{\mu_0,d,r})$ will depend on μ_0, r and d. In general, $\mathcal{I}_0$ and N_0 are increasing with r (w.r.t. $\subseteq$ and $\leq$, respectively), while Δ_0 will decrease as r increases between the jumps in N_0, but increase sharply at the jumps. Thus, the lexicographical ordering on (N_0, Δ_0^{-1}) coincides with increasing the radius, r. Since for nested constraint sets it should be more difficult to compete with the larger of the two constraints, it is intuitive that the lexicographical order on (N_0, Δ_0^{-1}) is an assessment of the difficulty of competing with a given constraint set.

EXAMPLE 6 (Convex hull of basic distributions). As motivated in Section 3.2, a natural setting is where $\mathcal{D}$ is the convex hull of some basic underlying distributions. Suppose $N = 3$, and there exist $\mu, \nu \in \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$ satisfying $\mu\ell = (0, 1, 0.5 + \varepsilon)$ and $\nu\ell = (1, 0, 0.5 + \varepsilon)$, where $\varepsilon > 0$ is arbitrary. Set $\mathcal{D} = \{\alpha\mu + (1 - \alpha)\nu \mid \alpha \in [0, 1]\}$, which gives $\mathcal{I}_0(\mathcal{D}) = \{1, 2\}$ and $\Delta_0(\mathcal{D}) = \varepsilon$.

However, on any given round it is possible for the data to be sampled from either μ or ν, in which case one of the effective experts is as separated (in expectation) as possible from the best expert and separated by an arbitrarily large multiplicative factor of Δ_0 from the ineffective expert. That is, this example demonstrates effective experts need not be better or even close to ineffective experts on any given round.

Note that Example 3 is related to the setting of Seldin and Slivkins [50] and Example 5 is related to the setting of Lykouris, Mirrokni and Paes Leme [41] (both focusing on bandit feedback), with the distinction that the existing literature considers constraints on the *cumulative* losses. In contrast, our constraints apply to the distributions allowed on any instantaneous round, and are not restricted in how they accumulate. This distinction is subtle, yet crucial to the type of adaptivity we propose in this work. While existing "easy data" results are about adapting to post-hoc summary statistics of the data, we provide adaptivity to the unknown, underlying dependence structure and propose that statistical methods should be designed to adapt to this as well (beyond adaptivity to model regularity assumptions).

4. Adaptive optimality for the semi-adversarial spectrum. In this section, we will state our main results that characterize the minimax regret over time-homogeneous convex constraints. We begin by precisely defining what it means for a prediction algorithm to be *adaptively minimax optimal*.

4.1. *Adaptively minimax optimal prediction algorithms.* Informally, an adaptively minimax optimal prediction algorithm achieves the minimax optimal regret (asymptotically in T) for the characterizing quantities constraining the allowable data-generating mechanism without *a priori* information on what values these characterizing quantities take. For collections of sequences $a = \{(a_{N,(N_0,\Delta_0)}(T))_{T\in\mathbb{N}} \mid N \in \mathbb{N}, (N_0, \Delta_0) \in [N] \times \mathbb{R}_+\}$ and $b = \{(b_{N,(N_0,\Delta_0)}(T))_{T\in\mathbb{N}} \mid N \in \mathbb{N}, (N_0, \Delta_0) \in [N] \times \mathbb{R}_+\}$, we write

$$a_{N,(N_0,\Delta_0)}(T) \lesssim b_{N,(N_0,\Delta_0)}(T) \quad \text{(abbreviated } a \lesssim b)$$

when

$$\begin{aligned} &\exists C > 0 \quad \forall N \in \mathbb{N}, \quad (N_0, \Delta_0) \in [N] \times \mathbb{R}_+ \quad \exists T_0 \in \mathbb{N} \quad \forall T > T_0, \\ &a_{N,(N_0,\Delta_0)}(T) \le C b_{N,(N_0,\Delta_0)}(T). \end{aligned} \tag{1}$$

If $a \lesssim b$ and $b \lesssim a$, we write $a_{N,(N_0,\Delta_0)}(T) \asymp b_{N,(N_0,\Delta_0)}(T)$ (abbreviated $a \asymp b$).

For a prediction algorithm $\mathfrak{a} = (\mathfrak{a}(N))_{N\in\mathbb{N}}$, we refer to equivalences class under $\asymp$ of

$$N, (N_0, \Delta_0), T \mapsto \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathfrak{a}} R(T)$$

as the *rate of regret* or simply the *rate* of $\mathfrak{a}$, and the equivalence class under $\asymp$ of

$$N, (N_0, \Delta_0), T \mapsto \inf_{\hat{\pi}\in\hat{\mathscr{P}}_N} \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\hat{\pi}} R(T)$$

as the *minimax optimal rate of regret*. Then we say a prediction algorithm $\mathfrak{a}$ is *adaptively minimax optimal* if

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathfrak{a}} R(T) \asymp \inf_{\hat{\pi}\in\hat{\mathscr{P}}_N} \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\hat{\pi}} R(T). \tag{2}$$

Further, we say that $\mathfrak{a}$ is *adaptive* if $\sup_{\pi\in\mathscr{P}_N} \mathbb{E}_{\pi,\mathfrak{a}} R(T)$ is always sublinear in T and, for some (N_0, Δ_0), its rate of regret is strictly better than the rate of $\inf_{\hat{\pi}\in\hat{\mathscr{P}}_N} \sup_{\pi\in\mathscr{P}_N} \mathbb{E}_{\pi,\hat{\pi}} R(T)$; otherwise, we say $\mathfrak{a}$ is *nonadaptive*. This definition formalizes the notion that an adaptive prediction algorithm must realize potential benefits from at least some instance of "easier" characterizing quantities and simultaneously have average regret at least converge to zero in all cases. Similar approaches to rule out trivial algorithms that achieve zero regret for one setting and linear regret in all other settings have been used for full-information [36] and bandits [38], Section 16.

Importantly, we do not demand that the prediction algorithm perform as well as if they had *a priori* knowledge of the true data-generating mechanism, since with this information the minimax regret can be quite small (zero or even negative). Instead, the prediction algorithm is only adapting to the *problem hardness*, as measured by the characterizing quantities, and consequently, there is still freedom in the minimax definition for the player to face its worst-case data-generating mechanism subject to these characterizing quantities. Mathematically, this is ensured by placing $\inf_{\hat{\pi}\in\hat{\mathscr{P}}_N}$ after the choice of characterizing quantities, but before the choice of the data-generating mechanism (i.e., $\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}}$).

More abstractly, our definition of adaptively minimax optimal can be interpreted under a generic adaptive decision problem, with a generic *problem size* given by N and a generic *problem hardness* replacing characterizing quantities. For example, in the case of density estimation, the problem size may correspond to the dimension of the data space, which the statistician knows, and the problem hardness may correspond to the Hölder continuity parameter of the true data-generating density, which the statistician does not know. For a further discussion of our definition of adaptively minimax optimal, see Section 4.3.

4.2. *Minimax rates*. We are now able to state our main result, establishing the minimax optimal rate of regret and that it is achieved by our novel algorithm META-CARE, which follows from the conjunction of Theorems 3 and 7 and Proposition 2.

THEOREM 1 (Main result).

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T) \asymp \inf_{\hat{\pi}\in\hat{\mathscr{P}}_N} \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\hat{\pi}} R(T) \asymp \sqrt{T\log N_0} + \frac{\log N}{\Delta_0}.$$

There is a natural interpretation of both terms in Theorem 1: $\sqrt{T\log N_0}$ is the minimax optimal rate of regret for adversarial losses with N_0 experts, while $(\log N)/\Delta_0$ is the minimax optimal rate of regret for stochastic losses with N experts and a stochastic gap of size Δ_0. Below, we elaborate on how we exploit intermediate results to obtain this optimal rate without knowledge of N_0 or Δ_0.

First, in Section 7 we investigate the behavior of HEDGE-like algorithms. In Theorem 4, we show that D.HEDGE using any parametrization that simultaneously achieves the minimax optimal rate of regret in both the stochastic-with-a-gap and adversarial settings is nonadaptive. That is, for $N_0 \geq 2$,

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \gtrsim \sqrt{T\log N}.$$

In fact, from Theorems 4 and 5, we find that without an *oracle parametrization* of D.HEDGE (one where N_0 is available to the player), it is only possible to achieve

$$\log(N_0)\sqrt{T} + \frac{(\log N)}{\Delta_0} \lesssim \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \lesssim \log(N_0)\sqrt{T} + \frac{(\log N)^2}{\Delta_0}$$

or

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \asymp \mathbb{I}_{[N_0\geq 2]}\sqrt{T\log N} + \frac{\log N}{\Delta_0},$$

but not both.

Then, in Section 8, we introduce another novel algorithm: FTRL-CARE. As an intermediary step that is critical for the proof of Theorem 1, we show in Theorem 6 that FTRL-CARE adapts with a better rate, satisfying

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{C}} R(T) \lesssim \sqrt{T\log N_0} + \frac{(\log N)^{3/2}}{\Delta_0}.$$

This bound achieves the minimax optimal rate for $N_0 \geq 2$, but not for $N_0 = 1$. To also guarantee the minimax optimal rate for $N_0 = 1$ (and consequently be adaptively minimax optimal), we introduce META-CARE in Theorem 7, which corresponds to another application of

D.HEDGE to the "meta-experts" corresponding to FTRL-CARE and D.HEDGE on all N experts; this proof relies heavily on our intermediate result for FTRL-CARE. We do not know whether the suboptimality of the bound for FTRL-CARE in the $N_0 = 1$ is a consequence of our analysis or is a true limitation of the method. Regardless, by invoking our meta-learning algorithm, META-CARE, we are able to guarantee adaptive minimax optimality in all cases.

Our quantitative upper bounds also explicitly demonstrate how large T must be for algorithms to have adaptive rates (i.e., expected regret that depends on N_0 and Δ_0), as opposed to the pessimistic adversarial rate (i.e., $\sqrt{T \log N}$). In particular, for both D.HEDGE and FTRL-CARE, roughly Δ_0^{-2} rounds of adversarial regret are incurred before the level of adaptation is sufficient to reduce the rate of regret accumulation. This demonstrates that as Δ_0 tends to 0, the player does not incur infinite regret from the Δ_0^{-1} terms, but rather incurs adversarial regret for a longer amount of time. We emphasize that the player does not need to know when they will stop incurring adversarial regret ahead of time to parametrize either algorithm, so knowledge of N_0 or Δ_0 is not required.

Our theoretical results are further supported by a simulation study that appears in Appendix H in the Supplementary Material [11]. The simulation study is based on the data-generating mechanisms that achieve the lower bound in the stochastic-with-a-gap setting and the algorithm specific lower bound for D.HEDGE with two effective experts. The results of the simulations agree with our theoretical results.

4.3. *Discussion on adaptive minimax optimality.* One might ask whether it is possible to strengthen the notion of adaptivity to be *uniform-in-T*, where the rate has to be achieved up to a constant at all T, rather than only for sufficiently large T depending on (N_0, Δ_0). This corresponds to replacing the relation $a \precsim b$ with the one defined by

$$\exists C > 0 \quad \forall T, \quad N \in \mathbb{N}, \quad (N_0, \Delta_0) \in [N] \times \mathbb{R}_+, \quad a_{N,(N_0,\Delta_0)}(T) \le C b_{N,(N_0,\Delta_0)}(T).$$

In the context of minimax regret, uniform adaptivity would require understanding the entire path of the regret (over T) rather than simply its eventual upper bound. This is not understood even in the stochastic setting; regret bounds of the form $1/\Delta$ in both the bandit and full-information settings (e.g., [6, 23, 44]) are all eventual upper bounds that are only known to be tight (i.e., have matching lower bounds) for sufficiently large T. Since identifying the minimax optimal regret uniformly in T remains open even for this basic setting, we do not attempt to solve this in our more general setting.

Beyond prediction with expert advice, the lack of uniform adaptivity also persists. For example, the leading constant of the minimax rates for smoothness-adaptation in statistics often depends on the smoothness parameter, which violates uniformity. For general questions of adaptive minimax optimality in sequential prediction, it is not clear how to demonstrate that either form of adaptivity is possible other than by constructing adaptive algorithms, as we have done in the present work.

Finally, one could consider adapting to a different collection of characterizing quantities than (N_0, Δ_0). For our setting, a natural extension is to consider the individual expectation gaps of each expert, rather than only the smallest gap. While our upper bounds can be extended to handle multiple gaps without much difficulty, tight lower bounds that depend on all the gaps simultaneously are again unknown even in the i.i.d. setting for full-information feedback. Since our work is about identifying minimax optimality, which would require such lower bounds, we do not consider this refinement. Beyond the extension to multiple gaps, it is an interesting avenue for future work to identify other characterizing quantities that could provide a finer characterization of the data-generating mechanism.

5. Concentration of measure for the semi-adversarial spectrum. In this section, we state and prove a concentration of measure result for data-generating mechanisms permitted by time-homogeneous convex constraints, which we use repeatedly to establish upper bounds on expected regret for D.HEDGE, FTRL-CARE and META-CARE. The result demonstrates that, even though the best expert may vary from round to round, the gap between the best effective expert along the observed data path and any ineffective expert grows like a sum of uniformly sub-Gaussian random variables with mean below $-\Delta_0$.

THEOREM 2. *For all $N \geq 2$, prediction policies $\hat{\pi} \in \hat{\mathscr{P}}_N$, convex sets $\mathcal{D} \subseteq \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$, $\lambda > 0$, $T_0 < T_1$ and $i \in [N] \setminus \mathcal{I}_0$,*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \min_{i_0 \in \mathcal{I}_0} \exp\left\{\lambda \sum_{t=T_0+1}^{T_1} \left[\ell_{i_0}(t) - \ell_i(t)\right]\right\} \leq \exp\{(T_1 - T_0)[\lambda^2/2 - \lambda\Delta_0]\}.$$

Note that we require the constraint set $\mathcal{D}$ to be convex. If $\mathcal{D}$ is not natively convex, our results clearly apply to its convex hull. There is, however, a natural reason to consider convex constraint sets: given a set $\mathcal{D}$ of joint distributions available for the data-generating mechanisms, requiring the set to be convex is equivalent to also allowing mixtures of the original available distributions. That is, the environment and experts together can randomly select a distribution from $\mathcal{D}$ to generate data from at each round.

One may wonder whether this result follows from an application of the Azuma–Hoeffding inequality. However, as demonstrated in Example 6, there exist simple constraint sets such that on any round, any effective expert (including the best overall) may have an arbitrarily larger expected loss than any ineffective expert. That is, $L_i(t) - L_{i_0}(t)$ need not be a (sub)martingale, and consequently, Azuma–Hoeffding does not directly apply. Instead, the proof of this result uses a variant of Sion's minimax theorem—which is the technical reason why we require the constraint set $\mathcal{D}$ to be convex—before applying Hoeffding's inequality to the instantaneous rounds. We restate the minimax theorem we require for completeness here, which can be found in Sion [51], Corollary 3.3, and has a simple proof in Komiya [33].

PROPOSITION 1. *Let $\mathcal{X}$ and $\mathcal{Y}$ be convex subsets of linear topological spaces, and suppose that $\mathcal{X}$ is compact. Let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be such that:*

(i) *for all $y \in \mathcal{Y}$, $f(\cdot, y) : \mathcal{X} \to \mathbb{R}$ is convex and upper semicontinuous; and*
(ii) *for all $x \in \mathcal{X}$, $f(x, \cdot) : \mathcal{Y} \to \mathbb{R}$ is concave and lower semicontinuous.*

Then $\inf_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} f(x, y) = \sup_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} f(x, y)$.

PROOF OF THEOREM 2. Let $\mathbb{R}_+ = [0, \infty)$, $\mathtt{simp}(\mathcal{I}_0) = \{v \in \mathbb{R}_+^{\mathcal{I}_0} : \sum_{i_0\in\mathcal{I}_0} v_{i_0} = 1\}$. First, since at least one optimal solution to a linear program on a compact convex polytope must be at a vertex,

$$\min_{i_0\in\mathcal{I}_0} \sum_{t=T_0+1}^{T_1} \left[\ell_{i_0}(t) - \ell_i(t)\right] = \inf_{v\in\mathtt{simp}(\mathcal{I}_0)} \sum_{t=T_0+1}^{T_1} \left[\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)\right].$$

Further, since exp is a monotone function, this identity implies

$$\min_{i_0\in\mathcal{I}_0} e^{\lambda \sum_{t=T_0+1}^{T_1}[\ell_{i_0}(t)-\ell_i(t)]} = \inf_{v\in\mathtt{simp}(\mathcal{I}_0)} e^{\lambda \sum_{t=T_0+1}^{T_1}[\langle v,\ell_{\mathcal{I}_0}(t)\rangle-\ell_i(t)]}.$$

Then, applying Jensen's and the max-min inequality give

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]} \\ \leq \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]}.$$

By the tower rule for conditional expectation and the definition of the kernel π_{T_1},

$$\mathbb{E}_{\pi,\hat{\pi}} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]} \\ \leq \left(\mathbb{E}_{\pi,\hat{\pi}}\left[e^{\lambda \sum_{t=T_0+1}^{T_1-1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]}\right]\right)\Big(\sup_{\mu \in \mathcal{D}} \mu\big(e^{\lambda[\langle v, \ell_{\mathcal{I}_0}\rangle - \ell_i]}\big)\Big).$$

Iterating this argument $T_1 - T_0 - 1$ more times, and using monotonicity of power functions, give

$$\inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]} \\ \leq \Big[\inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \sup_{\mu \in \mathcal{D}} \mu\big(e^{\lambda[\langle v, \ell_{\mathcal{I}_0}\rangle - \ell_i]}\big)\Big]^{T_1 - T_0}.$$

We now verify the conditions of Proposition 1. First, both $\mathtt{simp}(\mathcal{I}_0)$ and $\mathcal{D}$ are convex subsets of linear topological spaces, and $\mathtt{simp}(\mathcal{I}_0)$ is trivially compact (in fact, both are compact). Second, the objective function $f(v, \mu) = \mu(e^{\lambda[\langle v, \ell_{\mathcal{I}_0}\rangle - \ell_i]})$ is continuous and convex in v. Third, f is linear (and hence concave) in μ. Moreover, since each ℓ is bounded, for every v the objective f corresponds to integration of a bounded continuous function against a finite measure; hence, f is continuous with respect to μ in the topology of weak convergence. Thus, Proposition 1 gives

$$\inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \sup_{\mu \in \mathcal{D}} f(v, \mu) = \sup_{\mu \in \mathcal{D}} \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} f(v, \mu).$$

Thus,

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]} \\ \leq \Big[\sup_{\mu \in \mathcal{D}} \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \mu\big(e^{\lambda[\langle v, \ell_{\mathcal{I}_0}\rangle - \ell_i]}\big)\Big]^{T_1 - T_0}.$$

Consider any $\mu \in \mathcal{D}$, and let $i^*(\mu) \in \arg\min_{i \in [N]} \mu \ell_i$. By the definition of Δ_0, $\mu(\ell_{i^*(\mu)} - \ell_i) \leq -\Delta_0$ for every $i \in [N] \setminus \mathcal{I}_0$. Finally, since $\ell \in [0, 1]^N$ μ-a.s., by Hoeffding's lemma,

$$\inf_{v \in \mathtt{simp}(\mathcal{I}_0)} \mu\big(e^{\lambda[\langle v, \ell_{\mathcal{I}_0}\rangle - \ell_i]}\big) \leq \mu\big(e^{\lambda[\ell_{i^*(\mu)} - \ell_i]}\big) \leq e^{\lambda^2/2 - \lambda\Delta_0}.$$

Since this holds for all $\mu \in \mathcal{D}$,

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \inf_{v \in \mathtt{simp}(\mathcal{I}_0)} e^{\lambda \sum_{t=T_0+1}^{T_1} [\langle v, \ell_{\mathcal{I}_0}(t)\rangle - \ell_i(t)]} \leq \big[e^{\lambda^2/2 - \lambda\Delta_0}\big]^{T_1 - T_0} = e^{(T_1 - T_0)[\lambda^2/2 - \lambda\Delta_0]}.$$

□

6. Minimax lower bounds. In this section, we characterize the best possible performance under relaxations of the i.i.d. assumption. In particular, we quantify the best any prediction policy can do with oracle knowledge of the number of effective experts. The proof of this result is found in Appendix F.1 in the Supplementary Material [11]. While we do not expect a player to know the nature of the constraint set, we use this oracle lower bound to conclude that since our novel algorithm META-CARE achieves the same performance without using oracle knowledge, it is adaptively minimax optimal.

THEOREM 3. *There exist $\hat{\mathcal{Y}}$, $\mathcal{Y}$ and ℓ such that, for all $N_0 \in \mathbb{N}$, there exists $t_0 \in \mathbb{N}$ such that for all $N \in \mathbb{N}$ with $N \geq N_0$ and $T \geq t_0$,*

$$\sup_{\mathcal{D}\in\mathcal{V}(N,N_0,1/2)} \sup_{\pi\in\mathscr{P}(\mathcal{D})} \inf_{\hat{\pi}\in\mathscr{P}_N} \mathbb{E}_{\pi,\hat{\pi}} R(T) \geq \frac{\sqrt{T\log N_0}}{10}.$$

Theorem 3 allows us to characterize the minimax optimal dependence on T and N_0 of a prediction policy. However, for the case of $N_0 = 1$, the leading term instead depends on Δ_0. Consequently, to determine the minimax optimal rate of regret at all relaxations of the i.i.d. assumption, we must also use the following result by Mourtada and Gaïffas [44], which establishes a lower bound for when there is only one effective expert.

PROPOSITION 2 (Mourtada and Gaïffas [44], Proposition 4). *For all $N \in \mathbb{N}$, there exist $\hat{\mathcal{Y}}$, $\mathcal{Y}$ and ℓ such that for all $\Delta \in (0, 1/4)$ and $T \geq \frac{\log N}{16\Delta^2}$,*

$$\inf_{\hat{\pi}\in\mathscr{P}_N} \sup_{\mathcal{D}\in\mathcal{V}(N,1,\Delta)} \sup_{\pi\in\mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} R(T) \geq \frac{\log N}{256\Delta}.$$

These two lower bounds set the bar for what one should hope to achieve. In order to adapt to an unknown number of effective experts $N_0 \leq N$ and identity of the effective experts, the player can be forced to incur $\max(\sqrt{T\log N_0}, \Delta_0^{-1}\log N)$ rate of regret. Because $\max\{\sqrt{T\log N_0}, \Delta_0^{-1}\log N\} \asymp \sqrt{T\log N_0} + \Delta_0^{-1}\log N$, a prediction algorithm with a rate of regret $\lesssim \sqrt{T\log N_0} + \Delta_0^{-1}\log N$ is adaptively minimax optimal.

7. Performance of D.HEDGE. In this section, we show that without oracle knowledge of the characterizing quantities, D.HEDGE can be parametrized to either (a) be minimax optimal for the special case when $N_0 \in \{1, N\}$, but incur adversarial regret in between, or (b) adapt suboptimally to every value of the characterizing quantities. Following this section, we introduce FTRL-CARE and prove it adapts minimax optimally when there are multiple effective experts. We then *boost* these two algorithms together in META-CARE, and prove this is adaptively minimax optimal.

All of these prediction algorithms produce *proper prediction policies*, which means that rather than picking $\hat{y}$ from the entirety of $\hat{\mathcal{Y}}$, at each round the player chooses one of the experts $i \in [N]$ to emulate and predicts $\hat{y}(t) = x_i(t)$. To choose the expert to emulate, the history is used to choose a distribution on $[N]$, and then i is sampled from this distribution.

Formally, for $x \in \hat{\mathcal{Y}}^N$ and $w \in \texttt{simp}([N])$, let $x_\sharp w = \sum_{i\in[N]} w_i\delta_{x_i} \in \mathcal{M}(\hat{\mathcal{Y}})$ be the push-forward of $w \in \texttt{simp}([N])$ through x, viewing the vector x as a function $x : [N] \to \hat{\mathcal{Y}}$ and identifying $\texttt{simp}([N])$ with $\mathcal{M}([N])$. A proper prediction policy $\hat{\pi}^\star = (\hat{\pi}_t^\star)_{t\in\mathbb{N}}$ is any prediction policy such that, for all $t \in \mathbb{N}$, there exists a measurable map $w_t^\star : \mathcal{H}^{t-1} \to \texttt{simp}([N])$ satisfying, for all $h \in \mathcal{H}^{t-1}$ and $x \in \hat{\mathcal{Y}}^N$, $\hat{\pi}_t^\star((h, x), \cdot) = x_\sharp[w_t^\star(h)]$. The $\sigma(h(t-1))$-measurable random variable $w(t) = w_t^\star(h(t-1))$ is called the *weight vector*, or simply the *weights*. For each $i \in [N]$, $w_i(t)$ corresponds to the probability that the player will emulate the ith expert's prediction at time t.

The prediction algorithm HEDGE is parametrized by a sequence of measurable functions $(\tilde{\eta}_t)_{t\in\mathbb{N}} \in \prod_{t\in\mathbb{N}}\{\mathcal{H}^{t-1} \to \mathbb{R}_+\}$. The $\sigma(h(t-1))$-measurable random variable $\eta(t) = \tilde{\eta}_t(h(t-1))$ is called the *learning rate*, and the weights are defined by

$$w_i^{\mathrm{H}}(t) = \frac{\exp\{-\eta(t)L_i(t-1)\}}{\sum_{i'\in[N]}\exp\{-\eta(t)L_{i'}(t-1)\}}, \quad i \in [N].$$

The prediction algorithm DECREASING HEDGE (D.HEDGE) is parametrized by a function $g : \mathbb{N} \to \mathbb{R}_+$, and corresponds to HEDGE with the deterministic learning rate $\eta(t) = g(N)/\sqrt{t}$ for all $t \in \mathbb{N}$.

It is well known (see, e.g., Theorem 2.3 of [15]) that D.HEDGE with $g(N) \propto \sqrt{\log N}$ is minimax optimal in the adversarial setting, which corresponds to $\mathcal{D} = \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$. Recently, Mourtada and Gaïffas [44] showed that D.HEDGE with this parametrization is also minimax optimal in the i.i.d. setting, which corresponds to $|\mathcal{D}| = 1$. One might hope that this stochastic-and-adversarially minimax optimal parametrization would also perform well for all convex $\mathcal{D}$ in between these two cases. However, part (i) of Theorem 4 shows that, in fact, this parametrization fails to adapt to the number of effective experts when $N_0 \notin \{1, N\}$. Further, we show that a different parametrization can adapt in some ways, but does not achieve the minimax optimal dependence on T.

7.1. *Algorithm-specific lower bounds for* D.HEDGE. First, we observe that D.HEDGE with $g(N) \propto \sqrt{\log N}$, which is minimax optimal for both the stochastic and adversarial cases, does not adapt to an intermediate number of effective experts. Additionally, D.HEDGE with constant g can do better than the stochastic-and-adversarially minimax optimal parametrization, but cannot do as well as the oracle knowledge dependence on T from Theorem 3. We prove this in Appendix F.2 in the Supplementary Material [11].

Intuitively, to establish these lower bounds we use the same distribution as the adversarial lower bound, but restricted to a set of effective experts. This ensures that the expected regret of D.HEDGE is at least $\max\{g(N), H(\mathrm{Unif}(\mathcal{I}_0))/g(N)\} \cdot \sqrt{T}$, where H is the Shannon entropy. It follows that D.HEDGE cannot adapt because the parameter $g(N)$ would need to be set using information about the constraint set that is unavailable; namely, $H(\mathrm{Unif}(\mathcal{I}_0)) = \log N_0$.

THEOREM 4. (i) *For all* $c > 0$,

$$N \geq \exp\left\{\left(\frac{72\log 2}{c^2} + 9\right)e^{c^2/4}\right\} \quad \textit{and} \quad 2 \leq N_0 \leq e^{-c^2/8}N^{c^2\exp(c^2/4)/72} - 1,$$

there exist $\hat{\mathcal{Y}}$, $\mathcal{Y}$ *and* ℓ *such that for all* $T \geq 16c^{-2}\log N$, D.HEDGE *with* $g(N) = c\sqrt{\log N}$ *satisfies*

$$\sup_{\mathcal{D}\in\mathcal{V}(N,N_0,1/2)} \sup_{\pi\in\mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\mathrm{H}}R(T) \geq \frac{c\sqrt{T\log N}}{72\exp\{c^2/4\}} - \frac{1}{3c^2} - \frac{\log N}{3}.$$

(ii) *Suppose the player is allowed oracle knowledge of* N_0 *in addition to* N, *and consequently can parametrize* D.HEDGE *by any* $g : \mathbb{N}^2 \to \mathbb{R}_+$. *For all* $81 < N_0 \leq N$, *there exist* $\hat{\mathcal{Y}}$, $\mathcal{Y}$ *and* ℓ *such that* D.HEDGE *with* $g(N, N_0) \leq 2\sqrt{\log N_0 - 4\log 3}$ *satisfies that for all* $T \geq 32[g(N, N_0)]^{-2}\log N$,

$$\sup_{\mathcal{D}\in\mathcal{V}(N,N_0,1/2)} \sup_{\pi\in\mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\mathrm{H}}R(T) \geq \frac{\log(N_0)\sqrt{T}}{4g(N, N_0)} - \frac{3\log N_0}{[g(N, N_0)]^2}.$$

The proof of Theorem 4 can be used to argue that many existing "adaptive" variants of HEDGE will also fail to be adaptively minimax optimal along the semi-adversarial spectrum. We highlight this argument using well-known HEDGE variants from the literature. This is not meant to disparage these works, as they should not be expected to design algorithms for a notion of optimality defined years later, but to exemplify that adapting along the semi-adversarial spectrum is nontrivial and that the objectives of earlier works are insufficient to capture the notion of optimality we introduce.

The algorithm PROD of Cesa-Bianchi, Mansour and Stoltz [16] is essentially D.HEDGE with an adaptive learning rate shared by all experts. This adaptive learning rate is comprised of the reciprocal square-root of the cumulative squared losses, which will be (essentially) a constant multiple of t under the data-generating mechanism described in the proof of Theorem 4. Thus, the learning rate will behave the same as the data-independent learning rate of D.HEDGE, and consequently, a similar lower bound on performance applies. A similar argument would also hold for ADAHEDGE [21].

The refined algorithm ADAPT-ML-PROD of Gaillard, Stoltz and van Erven [23] is more subtle, since it has a different learning rate for each expert. However, the recommended learning rate (Corollary 4 of their paper) would not adapt optimally, because it uses $\log N$ for all experts, as opposed to an adaptive quantity as in FTRL-CARE. Consequently, for large enough N_0 and t, the data-generating mechanism of Theorem 4 will make the average loss with respect to the ADAPT-ML-PROD weights roughly $1/2$, and thus, ADAPT-ML-PROD inherits the same order of lower bound as D.HEDGE.

7.2. *Upper bounds for* D.HEDGE. Now, we show that the lower bound of Theorem 4 is tight. For a prediction policy $\hat{\pi}'$ that may be distinct from the actual prediction policy $\hat{\pi}$ the player is using, we define the *quasi-regret* (with respect to $\hat{\pi}'$) at time T by

$$\hat{R}_{\hat{\pi}'}(T) = \sum_{t=1}^{T} \int \ell(\hat{y}(t), y(t)) \hat{\pi}'_t\big((h(t-1), x(t)), \mathrm{d}\hat{y}(t)\big) - \min_{i \in [N]} \sum_{t=1}^{T} \ell(x_i(t), y(t)).$$

Quasi-regret replaces the actual loss at each round t with the conditional expectation of the player's loss *had that player played according to* $\hat{\pi}'_t$ *on round* t; the histories correspond, however, to the *actual* predictions made by $\hat{\pi}$. This allows us to quantify the performance of $\hat{\pi}'$ even when the entire sequence of predictions is governed by $\hat{\pi}$.

Clearly, $\mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\hat{\pi}}(T) = \mathbb{E}_{\pi,\hat{\pi}} R(T)$. However, we can prove almost sure results about $\hat{R}_{\hat{\pi}'}(T)$ for some prediction policy $\hat{\pi}'$, and then state expectation results of the form $\mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\hat{\pi}'}(T)$, where the expectation is with respect to a possibly different prediction policy $\hat{\pi}$. Results of this nature are crucial in the proof of Theorem 7, where we use them to control the regret accumulated by D.HEDGE and FTRL-CARE when the actual prediction policy is META-CARE.

THEOREM 5. *For all* $g : \mathbb{N} \to \mathbb{R}_+$ *used to parametrize* D.HEDGE, *all* $N \geq 2$, *prediction policies* $\hat{\pi} \in \hat{\mathscr{P}}_N$, *convex* $\mathcal{D} \subseteq \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$ *and* $T \in \mathbb{N}$,

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{H}}(T) \leq \sqrt{T+1}\left(\frac{\log N}{g(N)} + g(N)\right).$$

Moreover, when $T > \lceil \frac{8(\log N + [g(N)]^2/4 + g(N))^2}{[g(N)]^2 \Delta_0^2} \rceil$ *the following two cases hold*:

If $N_0 > 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{H}}(T) \leq \frac{17}{16}\sqrt{T}\left(\frac{\log N_0}{g(N)} + g(N)\right) + \frac{32}{\Delta_0}\left(\frac{\log N}{g(N)}\right)\left(\frac{\log N}{g(N)} + g(N)\right) + \sqrt{2}\left(\frac{\log N}{g(N)} + g(N)\right),$$

and if $N_0 = 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{H}}(T) \leq \frac{5}{\Delta_0}\left[\left(\frac{\log N}{g(N)}\right)\left(\frac{\log N}{g(N)} + g(N)\right) + 4\left(\frac{1}{g(N)^2} + g(N)^2\right)\right] + \sqrt{2}\left(\frac{\log N}{g(N)} + g(N)\right).$$

In order to more easily interpret this result, we also state the expected regret of D.HEDGE for various natural choices of g.

REMARK 1. Taking $\hat{\pi}$ to be determined by D.HEDGE:

(i) if $g(N)$ is constant, $\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \lesssim \log(N_0)\sqrt{T} + \frac{(\log N)^2}{\Delta_0}$;

(ii) if $g(N) \propto \sqrt{\log N}$, $\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \lesssim \mathbb{I}_{[N_0 \geq 2]}\sqrt{T \log N} + \frac{\log N}{\Delta_0}$;

(iii) in the oracle setting for $N_0 \geq 2$, if $g(N, N_0) \propto \sqrt{\log N_0}$,

$$\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{H}} R(T) \lesssim \sqrt{T \log N_0} + \frac{(\log N)^2}{\Delta_0 \log N_0}.$$

REMARK 2. If $g(N) \propto \sqrt{\log N}$, then Theorem 4(i) combined with Remark 1(ii) shows that the dependence on T is tight in Theorem 5. If oracle knowledge of N_0 is used to choose $g(N, N_0) \propto \sqrt{\log N_0}$, then Theorem 4(ii) simply matches the oracle lower bound of Theorem 3, confirming the dependence on T is tight in Theorem 5 (see Remark 1(iii)). Finally, if g is constant, then Theorem 4(ii) combined with Remark 1(i) shows that the dependence on T is tight in Theorem 5.

Together with the minimax lower bounds of Section 6, we find that, for the stochastic and adversarial settings, our expected regret bound for D.HEDGE with $g(N) = \sqrt{\log N}$ is tight up to constants and that the algorithm achieves the minimax optimal rates, as noted by Mourtada and Gaïffas [44]. Furthermore, we have improved upon Corollary 6 of Mourtada and Gaïffas [44] in the "adversarial-with-an-$\mathbb{E}$-gap" setting (see Example 4), having removed the extra $\Delta_0^{-1} \log(\Delta_0^{-1})$ dependence that separated the upper and lower bounds in their work.

8. Beating D.HEDGE without oracle knowledge. In Section 7, we completed the story of D.HEDGE by showing that it does not adapt minimax optimally to all possible constraint sets without oracle knowledge of the number of effective experts. It is natural to ask whether we can design an algorithm that adapts to the number of effective experts and has a rate of regret no larger than $\sqrt{T \log N_0}$.

In this section, we present a modified algorithm that does exactly this. Taking inspiration from the fact that D.HEDGE can be viewed as *follow-the-regularized-leader* (FTRL) using *entropic regularization* (see, e.g., Section 3.6 of [42]), we introduce the *constraint-adaptive root-entropic* (CARE) regularizer. We are able to prove upper bounds for the performance of FTRL for a large class of regularizers, and then use these upper bounds to prove both the upper bound results of Section 7 and the upper bounds for our improved algorithm, by viewing D.HEDGE and FTRL-CARE as FTRL with specifically chosen regularizers. Our bound shows that FTRL-CARE achieves the oracle rate $\sqrt{T \log N_0}$ without requiring knowledge of the characterizing quantities for the constraint set $\mathcal{D}$.

8.1. FTRL *algorithms*. FTRL is a generic method for online optimization. In the setting of sequential prediction with expert advice, FTRL is parametrized by a sequence of *regularizers* $\{r_t : \mathtt{simp}([N]) \to \mathbb{R}\}_{t\in\mathbb{Z}_+}$. Each such sequence, subject to regularity conditions on the regularizers (see Appendix C in the Supplementary Material [11]), determines a unique proper prediction policy. For each time $t+1$, a player using the $\text{FTRL}(\{r_t\}_{t\in\mathbb{Z}_+})$ algorithm has a proper prediction policy defined uniquely by the weight vectors given by

$$u(t+1) = \operatorname*{arg\,min}_{u\in\mathtt{simp}([N])} \big(\langle L(t), u\rangle + r_{0:t}(u)\big), \tag{3}$$

where $r_{0:t}(u) = \sum_{s=0}^{t} r_s(u)$, and the existence and uniqueness of the arg min is ensured by the regularity properties of the regularizer. This class of algorithms is well studied in online optimization; for specific results relevant to this work, see Appendix C in the Supplementary Material [11].

8.2. *The constraint-adaptive root-entropic regularizer.* First, we note that FTRL directly generalizes D.HEDGE. In particular, letting $H(u) = -\sum_{i\in[N]} u_i \log(u_i)$ denote the *entropy function*, it is well known that, for $r_{0:t}(u) = -\sqrt{t+1}H(u)/g(N)$, the weights played by $\text{FTRL}(\{r_t\}_{t\in\mathbb{Z}_+})$ are equal to the weights played by D.HEDGE. We modify the entropic regularizer to achieve improved performance for data-generating mechanisms strictly between stochastic and adversarial.

In order to motivate this new algorithm, we provide the following motivating intuition. First, from Remark 1, playing D.HEDGE with $g(N, N_0) \propto \sqrt{\log N_0}$ achieves the oracle rate. Second, the minimax optimal data-generating mechanism subject to the time-homogeneous convex constraint forces the minimax optimal prediction policy to "concentrate" to $\text{Unif}(\mathcal{I}_0)$. Finally, for $u = \text{Unif}(\mathcal{I}_0)$, $H(u) = \log N_0$. These three observations together suggest that, heuristically, playing HEDGE with the "adaptive" learning rate $\eta(t) = \sqrt{H(u(t))/t}$ may lead to an oracle rate of regret. However, $u(t)$ is defined in terms of $\eta(t)$, so this is an implicit system of equations to be solved at each time t. In order to define our modification of FTRL, we choose a regularizer such that the solution to the FTRL optimization problem gives rise to a similar system of equations. In particular, for some parameters $c_1, c_2 > 0$, the sequence of regularizers is given by

$$r_{0:t}(u) = -\frac{\sqrt{t+1}}{c_1}\sqrt{H(u) + c_2}. \tag{4}$$

We call $-r_0$ defined by equation (4) a *root-entropy function*, and regularization with $\{r_t\}_{t\in\mathbb{Z}_+}$ *constraint-adaptive root-entropic* (CARE) regularization. We refer to the algorithm $\text{FTRL}(\{r_t\}_{t\in\mathbb{Z}_+})$ with r_t induced by equation (4) as follow-the-regularized-leader with constraint-adaptive root-entropic regularization (or FTRL-CARE).

Throughout the remainder of the paper, we will use u for the weights output by the $\text{FTRL}(\{r_t\}_{t\in\mathbb{Z}_+})$ algorithm with a generic regularizer, w^{H} for weights output via entropic regularization (HEDGE) and w^{C} for weights output via root-entropic regularization (FTRL-CARE). The pseudocode for an efficient implementation of FTRL-CARE may be found in Appendix G in the Supplementary Material [11].

8.3. *Performance of* FTRL-CARE.

THEOREM 6. *For all $c_1, c_2 > 0$ used to parametrize* FTRL-CARE, *there exist $C_1, \ldots, C_4$ such that for all $N \geq 2$, prediction policies $\hat{\pi} \in \hat{\mathscr{P}}_N$, convex $\mathcal{D} \subseteq \mathcal{M}(\hat{\mathcal{Y}}^N \times \mathcal{Y})$ and $T \in \mathbb{N}$,*

$$\sup_{\pi\in\mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\text{C}}(T) \leq C_1\sqrt{(T+1)[\log N + c_2]}.$$

Moreover, when $T \geq \lceil \frac{2[\log N + C_4]^2}{c_1^2 c_2 \Delta_0^2} \rceil$, *the following two cases hold: if* $N_0 > 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{C}}(T) \leq \frac{33 C_1}{32} \sqrt{(T+1)[\log N_0 + c_2]} + C_2 \frac{[\log N + C_4]^{3/2}}{\Delta_0} + \frac{C_3}{\Delta_0},$$

and if $N_0 = 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{C}}(T) \leq C_2 \frac{[\log N + C_4]^{3/2}}{\Delta_0} + \frac{C_3 + 6}{\Delta_0}.$$

The constants $C_1, \dots, C_4$ *appearing above are given by*

$$C_1 = \left(\frac{1}{c_1} + \frac{3c_1}{2}\right), \qquad C_2 = \sqrt{2} C_1 \left(\frac{1}{c_1 \sqrt{c_2}} + \frac{1}{c_2}\right),$$

$$C_3 = \sqrt{2} \frac{8 + 12 c_1^2}{3 c_1^2 \sqrt{c_2}} \quad \text{and} \quad C_4 = \max\left\{c_2, 3c_1\sqrt{c_2} + \frac{5 c_1^2 c_2}{4}\right\}.$$

With $c_1 = c_2 = 1$, *this simplifies to: for all* $T \in \mathbb{N}$,

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{C}}(T) \leq 3\sqrt{(T+1)[\log N + 1]},$$

and when $T \geq \lceil \frac{2[\log N + 5]^2}{\Delta_0^2} \rceil$, *if* $N_0 > 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{C}}(T) \leq 3\sqrt{(T+1)[\log N_0 + 1]} + 8 \frac{[\log N + 5]^{3/2}}{\Delta_0} + \frac{10}{\Delta_0},$$

and if $N_0 = 1$, *then*

$$\sup_{\pi \in \mathscr{P}(\mathcal{D})} \mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{C}}(T) \leq 8 \frac{[\log N + 5]^{3/2}}{\Delta_0} + \frac{16}{\Delta_0}.$$

REMARK 3. Taking $\hat{\pi}$ to be determined by FTRL-CARE,

$$\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{C}} R(T) \lesssim \sqrt{T \log N_0} + \frac{(\log N)^{3/2}}{\Delta_0}.$$

REMARK 4. Note that in the case $N_0 = 1$, this is worse than D.HEDGE with learning rate $g(N) \propto \sqrt{\log N}$, which has $\Delta_0^{-1} \log N$ rate of regret. We resolve this in Section 9 by introducing a new algorithm, META-CARE, that combines the optimality of D.HEDGE in the stochastic case and FTRL-CARE elsewhere.

9. CARE if you can, HEDGE if you must or META-CARE for all. Since we have seen in Theorem 5 that D.HEDGE with $g(N) = \sqrt{\log N}$ achieves the minimax optimal order of $\log N$ when $N_0 = 1$, and Theorem 6 shows that FTRL-CARE is minimax optimal in all other cases, it is natural to try to combine these two algorithms in order to have minimax optimal rate of regret for all values of N_0 and Δ_0. To achieve this, we introduce the META-CARE algorithm.

Intuitively, META-CARE plays both D.HEDGE and FTRL-CARE, treating them as two *meta-experts*. META-CARE outputs the weighted average of the predictions made by the two meta-experts, where the weighting output by D.HEDGE based on their respective losses. Consequently, META-CARE has four parameters: $c_{\mathrm{H}}, c_{\mathrm{C},1}, c_{\mathrm{C},2}, c_{\mathrm{M}} > 0$. Formally, for each $t \in \mathbb{N}$, let $w^{\mathrm{H}}(t)$ denote the weight vector produced by D.HEDGE with $g(N) = c_{\mathrm{H}}\sqrt{\log N}$ at time

t and let $w^{\mathrm{C}}(t)$ denote the weight produced by FTRL-CARE with parameters $c_{\mathrm{C},1}$, $c_{\mathrm{C},2}$ at time t. Consider the *meta-losses* defined by

$$\ell_{\mathrm{H}}(t) = \langle \ell(t), w^{\mathrm{H}}(t) \rangle, \qquad \ell_{\mathrm{C}}(t) = \langle \ell(t), w^{\mathrm{C}}(t) \rangle,$$

$$L_{\mathrm{H}}(t) = \sum_{s=1}^{t} \ell_{\mathrm{H}}(t), \qquad L_{\mathrm{C}}(t) = \sum_{s=1}^{t} \ell_{\mathrm{C}}(t).$$

Then, for each $t \in \mathbb{N}$, META-CARE produces the weight vector

$$w^{\mathrm{M}}(t+1) = \frac{\exp\{-\eta_{\mathrm{M}}(t) L_{\mathrm{H}}(t)\} w^{\mathrm{H}}(t+1) + \exp\{-\eta_{\mathrm{M}}(t) L_{\mathrm{C}}(t)\} w^{\mathrm{C}}(t+1)}{\exp\{-\eta_{\mathrm{M}}(t) L_{\mathrm{H}}(t)\} + \exp\{-\eta_{\mathrm{M}}(t) L_{\mathrm{C}}(t)\}},$$

where $\eta_{\mathrm{M}}(t) = c_{\mathrm{M}}/\sqrt{t}$. Observe that $w^{\mathrm{M}}(t+1)$ will be an element of $\mathtt{simp}([N])$ since it is a convex combination of $w^{\mathrm{H}}(t+1)$ and $w^{\mathrm{C}}(t+1)$, both of which are elements of $\mathtt{simp}([N])$.

THEOREM 7. META-CARE *parametrized by* $c_{\mathrm{H}} = \sqrt{\log N}$ *and* $c_{\mathrm{C},1} = c_{\mathrm{C},2} = c_{\mathrm{M}} = 1$ *incurs*

$$\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T) \lesssim \sqrt{T \log N_0} + \frac{\log N}{\Delta_0}.$$

We do not state a detailed quantitative form of Theorem 7, since our proof can be easily extended for any arbitrary $\hat{\pi}$ to a bound on $\mathbb{E}_{\pi,\hat{\pi}} \hat{R}_{\mathrm{M}}(T)$ with exact constants using the statements and proofs of Theorems 5 and 6.

PROOF OF THEOREM 7. For $N_0 \geq 2$, we decompose the quasi-regret of META-CARE into components coming from the quasi-regret due to meta-learning and the quasi-regret of the better of the two meta-experts. In particular, for any sequence of losses $(\ell(t))_{t \in \mathbb{N}}$, we can write

$$\begin{aligned} \hat{R}_{\mathrm{M}}(T) &= \sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{M}}(t) \rangle - \min_{i \in [N]} \sum_{t=1}^{T} \ell_i(t) \\ &= \left[\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{M}}(t) \rangle - \min\left(\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{H}}(t) \rangle, \sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{C}}(t) \rangle \right) \right] \\ &\quad + \min(\hat{R}_{\mathrm{H}}(T), \hat{R}_{\mathrm{C}}(T)). \end{aligned}$$

Therefore, for any $N_0 \leq N$ and Δ_0,

$$\begin{aligned} &\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T) \\ &\quad \leq \sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \left[\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{M}}(t) \rangle - \min\left(\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{H}}(t) \rangle, \sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{C}}(t) \rangle \right) \right] \\ &\qquad + \sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \min(\hat{R}_{\mathrm{H}}(T), \hat{R}_{\mathrm{C}}(T)). \end{aligned}$$

First, we consider the case when $N_0 \geq 2$. Since META-CARE is D.HEDGE with two experts given by the predictions of D.HEDGE and FTRL-CARE, Theorem 5 implies that

$$\begin{aligned} &\sup_{\pi \in \mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \left[\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{M}}(t) \rangle - \min\left(\sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{H}}(t) \rangle, \sum_{t=1}^{T} \langle \ell(t), w^{\mathrm{C}}(t) \rangle \right) \right] \\ &\quad \leq \sqrt{T+1} \left(\frac{\log(2)}{c_{\mathrm{M}}} + \frac{3 c_{\mathrm{M}}}{4} \right). \end{aligned} \tag{5}$$

Then, since $(\log N)^{3/2}\Delta_0^{-1}$ is lower order according to our $\lesssim$ notation when $N_0 \geq 2$, from Theorem 6 we obtain

$$\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \min\bigl(\hat{R}_{\mathrm{H}}(T), \hat{R}_{\mathrm{C}}(T)\bigr) \leq \sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \hat{R}_{\mathrm{C}}(T) \lesssim \sqrt{T\log N_0}. \tag{6}$$

Combining equations (5) and (6) imply that, when $N_0 \geq 2$, $\sup_{\pi\in\mathscr{P}_{N,(N_0,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T) \lesssim \sqrt{T\log N_0}$. Now consider the case where $N_0 = 1$, and let $\mathcal{I}_0 = \{i_0\}$. Using a similar decomposition to the previous case, we have

$$\begin{aligned}\hat{R}_{\mathrm{M}}(T) &= \sum_{t=1}^{T}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \min_{i\in[N]}\sum_{t=1}^{T}\ell_i(t)\\ &= \Biggl[\sum_{t=1}^{T}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \sum_{t=1}^{T}\langle \ell(t), w^{\mathrm{H}}(t)\rangle\Biggr] + \hat{R}_{\mathrm{H}}(T).\end{aligned}$$

Let t_0 be as in Theorem 6 (with $(c_1, c_2) = (c_{\mathrm{C},1}, c_{\mathrm{C},2})$), so that $t_0 \lesssim \frac{(\log N)^2}{\Delta_0^2}$. Expanding the quasi-regret of META-CARE and using the boundedness of the losses give

$$\begin{aligned}\hat{R}_{\mathrm{M}}(T) &= \Biggl[\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{H}}(t)\rangle\Biggr]\\ &\quad + \Biggl[\sum_{t=t_0+1}^{T}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \sum_{t=t_0+1}^{T}\langle \ell(t), w^{\mathrm{H}}(t)\rangle\Biggr] + \hat{R}_{\mathrm{H}}(T)\\ &\leq \Biggl[\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \min\Biggl(\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{H}}(t)\rangle, \sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{C}}(t)\rangle\Biggr)\Biggr]\\ &\quad + \sum_{t=t_0+1}^{T}\frac{1}{2}\bigl\|w^{\mathrm{C}}(t) - w^{\mathrm{H}}(t)\bigr\|_{L^1} + \hat{R}_{\mathrm{H}}(T).\end{aligned}$$

Therefore,

$$\begin{aligned}&\sup_{\pi\in\mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T)\\ &\quad\leq \sup_{\pi\in\mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}}\Biggl[\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \min\Biggl(\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{H}}(t)\rangle, \sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{C}}(t)\rangle\Biggr)\Biggr]\\ &\qquad + \sup_{\pi\in\mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}}\sum_{t=t_0}^{\infty}\frac{1}{2}\bigl\|w^{\mathrm{C}}(t) - w^{\mathrm{H}}(t)\bigr\|_{L^1} + \sup_{\pi\in\mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}}\hat{R}_{\mathrm{H}}(T).\end{aligned}$$

Again using the fact that META-CARE is D.HEDGE with two experts given by the predictions of D.HEDGE and FTRL-CARE, by Theorem 5 we have

$$\begin{aligned}&\sup_{\pi\in\mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}}\Biggl[\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{M}}(t)\rangle - \min\Biggl(\sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{H}}(t)\rangle, \sum_{t=1}^{t_0}\langle \ell(t), w^{\mathrm{C}}(t)\rangle\Biggr)\Biggr]\\ &\quad\lesssim \sqrt{t_0} \lesssim \frac{\log N}{\Delta_0}.\end{aligned} \tag{7}$$

Next, using the triangle inequality, the fact that $1 - w^{\mathrm{H}}_{i_0}(t) = \sum_{i\in[N]\setminus\mathcal{I}_0} w^{\mathrm{H}}_i(t)$ along with the same fact for w^{C}, and Lemmas A.2 and E.3 (see also the proofs of Theorems 5 and 6 for more details),

$$\sup_{\pi \in \mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \sum_{t=t_0}^{\infty} \frac{1}{2} \| w^{\mathrm{C}}(t) - w^{\mathrm{H}}(t) \|_{L^1}$$

$$\leq \sup_{\pi \in \mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \sum_{t=t_0}^{\infty} \frac{1}{2} \big(\| w^{\mathrm{C}}(t) - \delta_{i_0} \|_{L^1} + \| w^{\mathrm{H}}(t) - \delta_{i_0} \|_{L^1} \big) \tag{8}$$

$$= \sup_{\pi \in \mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \sum_{t=t_0}^{T} \sum_{i \in [N] \setminus \mathcal{I}_0} \big(w_i^{\mathrm{H}}(t) + w_i^{\mathrm{C}}(t) \big) \lesssim \frac{1}{\Delta_0},$$

where δ_{i_0} is the point-mass on i_0 (equivalently, the weight vector with weight 1 on expert i_0 and 0 on the others).

Finally, from Theorem 5, $\sup_{\pi \in \mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} \hat{R}_{\mathrm{H}}(T) \lesssim \frac{\log N}{\Delta_0}$. Combining this with equations (7) and (8) shows that, in the case of $N_0 = 1$,

$$\sup_{\pi \in \mathscr{P}_{N,(1,\Delta_0)}} \mathbb{E}_{\pi,\mathrm{M}} R(T) \lesssim \frac{\log N}{\Delta_0}. \qquad \square$$

10. Related work. The existing literature on statistical decision making with sequential data is vast, spanning decades and at least two major fields of study: sequential decision theory began as a subfield of statistics, and the historical literature is rather exclusive to statistics, while the more recent literature on decision procedures without i.i.d. assumptions has largely been developed within machine learning and computer science. In this section, we highlight the most relevant notions of adaptivity, and how their statistical interpretations differ from each other as well as the present work.

10.1. *Distributional assumptions*. First, note that while we use the language of prediction to describe our setting, our prediction space $\hat{\mathcal{Y}}$ is distinct from the observation space $\mathcal{Y}$, so we achieve the same level of generality as allowing for arbitrary decisions. Classically, the statistical literature on sequential hypothesis testing [10, 18, 37, 49, 58] and sequential parameter estimation [2, 24, 48, 60] rely on assumptions on the joint dependence structure of data to obtain performance guarantees. From a minimax perspective, removing the assumptions on the dependence structure reduces the problem to adversarially chosen data. Instead, by characterizing these arbitrary distributions in some way such that performance depends on the characterization, we can design methods for which the performance adapts to the characterization.

Hanneke [28] provides an overview of when classical estimation procedures designed for i.i.d. data will be consistent under various nonstationarity conditions. Additionally, he considers the asymptotic performance of a broader class of algorithms, although there is no notion of *adaptivity* since performance is binary: either a dependence structure admits a consistent online learning algorithm or it does not. In contrast, since the present work deals with finite expert classes, there is always a consistent algorithm, and so we focus on the specific performance of algorithms beyond their convergence properties.

Rakhlin, Sridharan and Tewari [46] consider general constraints on the data-generating mechanism for sequential prediction. We also use constraints on the data-generating mechanism to define relaxations of the i.i.d. assumption, but the specific constraints that we define and study are not ones studied by Rakhlin, Sridharan and Tewari. Additionally, we focus on developing methods that are minimax optimal under the constraint even when the nature of the constraint is unknown. For each of the constraints analyzed by Rakhlin, Sridharan and Tewari, the authors bound the minimax regret nonconstructively, and consequently, cannot guarantee the existence of an algorithm that is adaptively minimax optimal. In contrast, we provide an explicit, efficient algorithm that is adaptively minimax optimal for our constraint framework.

10.2. *Notions of easy data.* Beyond quantifying the minimax performance of decision rules under distributional assumptions, significant progress has been made over the last decade toward regret bounds that depend on key summary statistics of the observed data sequence. While the terminology for these types of bounds varies in the literature, we will follow the nomenclature of Cesa-Bianchi, Mansour and Stoltz [16], who differentiate between zero-order, first-order and second-order regret bounds. We use stochastic constraints to link zero- and second-order bounds in a general framework, and hence can compare with results derived in a wide range of settings.

Zero-order bounds refer to those that depend only on the time horizon, the size of the expert class and an absolute bound on the size of the predictions (alternatively, the losses). Results of this nature have existed for many years, beginning with Littlestone and Warmuth [39] and Vovk [57], and are concisely summarized by Cesa-Bianchi and Lugosi [15]. These bounds are often dubbed *worst case* or *adversarial*, since they hold for any sequence of observations subject to the aforementioned global constraints.

In contrast, first-order bounds control regret in terms of a data-dependent quantity; namely, the sum of the actual observed losses (potentially over all experts, or just the best expert for tighter results). Hence, they may lead to much tighter bounds than zero-order guarantees if the observed losses end up being in a much tighter range than is guaranteed by some absolute bound on the size of the losses. The first bound of this form was by Freund and Schapire [22] for the HEDGE algorithm, which was later upgraded to a multiplicative rather than additive dependence on the cumulative best loss [15], Corollary 2.4. Similar bounds have been developed for the bandit setting [5, 7], algorithms with adaptive parametrization [31, 55] and the combination of adaptive parametrization with partial information [45].

However, a limitation of first-order bounds is that they are not translation-invariant in the losses. In particular, they suggest that every expert incurring loss of 1 on each round is much harder to compete against than every expert incurring loss of zero on each round, which is not the case. One solution is to obtain regret bounds that are similar to first order, but rather than depending on the sum of the losses, they depend on a single first-order translation-invariant parameter that characterizes the observed loss sequence. In the bandit setting, examples of such a parameter include the effective loss range [17, 52] and the amount of corruption allowed on the mean of the losses [26, 41]. A similar analysis of corruption of experts' predictions in the full-information setting has recently appeared by Amir et al. [1].

Beyond these first-order quantities, another line of work has focused on second-order bounds, which depend on some form of variation of the observed losses. The first results of this form were derived by Cesa-Bianchi, Mansour and Stoltz [16], who obtain a bound in terms of the sum of the squared losses via tuning the learning rate for D.HEDGE. This was extended by both McMahan and Streeter [43] and Hazan and Kale [29] to depend on the sample second moment and variance, respectively, of the losses (empirically along the trajectory of observations), and again by Hazan and Kale [30] to obtain the same in the bandit setting. Both Erven et al. [55] and de Rooij et al. [21] obtain similar variation bounds, which are smaller for a different notion of "easy" data (defined by the mixability of the loss). Finally, another type of second-order bound was developed by Gaillard, Stoltz and van Erven [23], where they utilize the squared difference of algorithm losses with expert losses.

A different perspective on easy data is taken by Chaudhuri, Freund and Hsu [19] and Luo and Schapire [40], who develop methods not only to have regret relative to the best expert of size $\mathcal{O}(\sqrt{T \log N})$, but to also have regret relative to the εN-quantile expert of size $\mathcal{O}(\sqrt{T\log(1/\varepsilon)})$ for all $\varepsilon \in (1/N, 1)$. The algorithms they propose are more optimistic than D.HEDGE in the sense that they trust the past data more, which leads to suboptimal performance in settings between stochastic and adversarial, exaggerating the shortcomings of the standard parametrization of D.HEDGE in this case.

Several other methods exist that tune the learning rate of HEDGE adaptively based on the past interaction with the environment. Generally, these are motivated by improved second-order bounds. Examples include Koolen and van Erven [35] and van Erven and Koolen [56], who use a prior on the learning rate and meta-experts for a discrete collection of possible learning rates, respectively.

We also derive second-order (in particular, variance) bounds for the observed data sequence (see the intermediary result Theorem A.1). However, we are also able to extend this notion due to the stochastic nature of our constraints. In particular, once we take the expectation (with respect to the data-generating mechanism and the player's actions) of our second-order bounds, we obtain bounds directly comparable to (and tighter than) existing zero-order bounds. This provides greater insight than existing second-order bounds, which often leave a direct dependence on the variability of the chosen learning algorithm that is not *a priori* clear, and do not explicitly characterize what an "easy" data sequence actually looks like.

In the full-information setting, another line of investigation describes "easy" stochastic data by that which satisfies a *Bernstein condition*; that is, the conditional second moment of the losses are controlled by a concave function of the conditional first moment. This condition was shown to be crucial for achieving *fast rates* in the batch setting by Bartlett and Mendelson [9], then in the online convex optimization setting (infinite expert class) by van Erven and Koolen [56] and finally for simultaneously the finite expert and infinite expert online setting by Koolen, Grünwald and van Erven [34]. Recent work by Grünwald and Mehta [25] provided sufficient conditions to extend these results to unbounded losses.

10.3. *Stochastic and adversarially optimal algorithms*. In addition to developing bounds for "easy" data, the line of work most relevant to the present paper has focused on developing algorithms that are simultaneously optimal in two key settings: worst-case adversarial observations and i.i.d. (stochastic) observations. These bounds are characterized by matching the adversarial bounds mentioned above and the optimal stochastic bounds for either bandits [6], Theorem 1, or full information [23], Theorem 11. Beginning with Audibert and Bubeck [4] and Bubeck and Slivkins [12], the bandit literature is rich in this area; contributions include removing prior knowledge of the time horizon [50], matching lower bounds [8] and a simultaneously optimal algorithm with respect to a slightly weaker notion of regret [62].

In our discussion of the previous bounds, we have not specifically distinguished between the types of algorithms used to achieve them. However, there is an aesthetic (and computational) desire to find algorithms that achieve regret bounds that are optimal both for worst-case data and some notion of "easy" data, and yet are as simple as the algorithms which perform well in either just the adversarial or just the i.i.d. setting. A recent breakthrough on this front was achieved by Mourtada and Gaïffas [44], who showed the standard parametrization of the D.HEDGE algorithm is optimal for both the adversarial and the stochastic settings. For the bandit setting, the $\frac{1}{2}$-TSALLIS-INF algorithm of Zimmert and Seldin [62] has a similarly simple aesthetic; namely, it is also an analytic solution to an FTRL problem with an appropriate regularizer. One of the more surprising contributions of our work is that we show *every* prespecified parametrization of D.HEDGE is not adaptively minimax optimal.

Acknowledgments. BB and JN are equal-contribution authors; order was determined randomly.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force. DR is also a faculty member of the Vector Institute. BB was a student affiliate at the Vector Institute while this work was undertaken. JN was a student affiliate at the Vector Institute when this work began, and is now a faculty affiliate there. This research was partially carried out while all three authors were visiting the Institute for Advanced Study in Princeton, New Jersey, for the Special Year on Optimization, Statistics and Theoretical Machine Learning. JN and BB's travel to the Institute for Advanced Study were separately funded by NSERC Michael Smith Foreign Study Supplements. We thank Nicolò Campolongo, Peter D. Grünwald, Teodor Vanislavov Marinov, Francesco Orabona, Alex Stringer, Csaba Szepesvári, Yanbo Tang and Julian Zimmert for their insightful comments on preliminary versions of this work.

Funding. BB was supported by an NSERC Canada Graduate Scholarship and the Vector Institute. JN was supported by an NSERC Vanier Canada Graduate Scholarship and the Vector Institute. DMR is supported in part by an NSERC Discovery Grant, Ontario Early Researcher Award, Canada CIFAR AI Chair funding through the Vector Institute, and a stipend provided by the Charles Simonyi Endowment. This material is based also upon work supported by the United States Air Force under Contract No. FA850-19-C-0511.

SUPPLEMENTARY MATERIAL

Supplement to: "Relaxing the i.i.d. assumption: Adaptively minimax optimal regret via root-entropic regularization" (DOI: 10.1214/23-AOS2315SUPP; .pdf). Additional proofs, additional proof details, implementation details, and simulations.

REFERENCES

[1] AMIR, I., ATTIAS, I., KOREN, T., LIVNI, R. and MANSOUR, Y. (2020). Prediction with corrupted expert advice. In *Advances in Neural Information Processing Systems* 33.

[2] ANSCOMBE, F. J. (1953). Sequential estimation. *J. Roy. Statist. Soc. Ser. B* **15** 1–29. MR0056895

[3] AUDIBERT, J.-Y. (2009). Fast learning rates in statistical inference through aggregation. *Ann. Statist.* **37** 1591–1646. MR2533466 https://doi.org/10.1214/08-AOS623

[4] AUDIBERT, J.-Y. and BUBECK, S. (2009). Minimax policies for adversarial and stochastic bandits. In *Proceedings of the* 22*nd Conference on Learning Theory*.

[5] AUDIBERT, J.-Y. and BUBECK, S. (2010). Regret bounds and minimax policies under partial monitoring. *J. Mach. Learn. Res.* **11** 2785–2836. MR2738783

[6] AUER, P., CESA-BIANCHI, N. and FISCHER, P. (2002). Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47** 235–256.

[7] AUER, P., CESA-BIANCHI, N., FREUND, Y. and SCHAPIRE, R. E. (2003). The nonstochastic multiarmed bandit problem. *SIAM J. Comput.* **32** 48–77. MR1954855 https://doi.org/10.1137/S0097539701398375

[8] AUER, P. and CHIANG, C.-K. (2016). An algorithm with nearly optimal pseudo-regret for both stochastic and adversarial bandits. In *Proceedings of the* 29*th Conference on Learning Theory*.

[9] BARTLETT, P. L. and MENDELSON, S. (2006). Empirical minimization. *Probab. Theory Related Fields* **135** 311–334. MR2240689 https://doi.org/10.1007/s00440-005-0462-3

[10] BEGG, C. B. and MEHTA, C. R. (1979). Sequential analysis of comparative clinical trials. *Biometrika* **66** 97–103. MR0529153 https://doi.org/10.1093/biomet/66.1.97

[11] BILODEAU, B., NEGREA, J. and ROY, D. M. (2023). Supplement to "Relaxing the i.i.d. assumption: Adaptively minimax optimal regret via root-entropic regularization." https://doi.org/10.1214/23-AOS2315SUPP

[12] BUBECK, S. and SLIVKINS, A. (2012). The best of both worlds: Stochastic and adversarial bandits. In *Proceedings of the* 25*th Conference on Learning Theory*.

[13] CAI, T. T. and LOW, M. G. (2004). An adaptation theory for nonparametric confidence intervals. *Ann. Statist.* **32** 1805–1840. MR2102494 https://doi.org/10.1214/009053604000000049

[14] CESA-BIANCHI, N. and LUGOSI, G. (1999). On prediction of individual sequences. *Ann. Statist.* **27** 1865–1895. MR1765620 https://doi.org/10.1214/aos/1017939242

[15] CESA-BIANCHI, N. and LUGOSI, G. (2006). *Prediction, Learning, and Games*. Cambridge Univ. Press, Cambridge. MR2409394 https://doi.org/10.1017/CBO9780511546921

[16] CESA-BIANCHI, N., MANSOUR, Y. and STOLTZ, G. (2007). Improved second-order bounds for prediction with expert advice. *Mach. Learn.* **66** 321–352.

[17] CESA-BIANCHI, N. and SHAMIR, O. (2018). Bandit regret scaling with the effective loss range. In *Algorithmic Learning Theory* 2018. *Proc. Mach. Learn. Res.* (*PMLR*) **83** 24. Proceedings of Machine Learning Research PMLR. MR3857304

[18] CHAMBAZ, A., ZHENG, W. and VAN DER LAAN, M. J. (2017). Targeted sequential design for targeted learning inference of the optimal treatment rule and its mean reward. *Ann. Statist.* **45** 2537–2564. MR3737901 https://doi.org/10.1214/16-AOS1534

[19] CHAUDHURI, K., FREUND, Y. and HSU, D. J. (2009). A parameter-free hedging algorithm. In *Advances in Neural Information Processing Systems* 22.

[20] COVER, T. M. (1967). Behavior of sequential predictors of binary sequences. In *Trans. Fourth Prague Conf. on Information Theory, Statistical Decision Functions, Random Processes* (*Prague*, 1965) 263–272. Academia, Prague. MR0217944

[21] DE ROOIJ, S., VAN ERVEN, T., GRÜNWALD, P. D. and KOOLEN, W. M. (2014). Follow the leader if you can, hedge if you must. *J. Mach. Learn. Res.* **15** 1281–1316. MR3214784

[22] FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139. MR1473055 https://doi.org/10.1006/jcss.1997.1504

[23] GAILLARD, P., STOLTZ, G. and VAN ERVEN, T. (2014). A second-order bound with excess losses. In *Proceedings of the* 27*th Conference on Learning Theory*.

[24] GILLILAND, D. C. (1968). Sequential compound estimation. *Ann. Math. Stat.* **39** 1890–1904. MR0285067 https://doi.org/10.1214/aoms/1177698019

[25] GRÜNWALD, P. D. and MEHTA, N. A. (2020). Fast rates for general unbounded loss functions: From ERM to generalized Bayes. *J. Mach. Learn. Res.* **21** Paper No. 56, 80 pp. MR4095335

[26] GUPTA, A., KOREN, T. and TALWAR, K. (2019). Better algorithms for stochastic bandits with adversarial corruptions. In *Proceedings of the* 32*nd Conference on Learning Theory*.

[27] HANNAN, J. (1957). Approximation to Bayes risk in repeated play. In *Contributions to the Theory of Games, Vol.* 3. *Ann. of Math. Stud.* **39** 97–139. Princeton Univ. Press, Princeton, NJ. MR0109094

[28] HANNEKE, S. (2017). Learning whenever learning is possible: Universal learning under general stochastic processes. Preprint. Available at arXiv:1706.01418.

[29] HAZAN, E. and KALE, S. (2010). Extracting certainty from uncertainty: Regret bounded by variation in costs. *Mach. Learn.* **80** 165–188. MR3108164 https://doi.org/10.1007/s10994-010-5175-x

[30] HAZAN, E. and KALE, S. (2011). Better algorithms for benign bandits. *J. Mach. Learn. Res.* **12** 1287–1311. MR2804601

[31] HUTTER, M. and POLAND, J. (2004). Prediction with expert advice by following the perturbed leader for general weights. In *Algorithmic Learning Theory*. *Lecture Notes in Computer Science* **3244** 279–293. Springer, Berlin. MR2167508 https://doi.org/10.1007/978-3-540-30215-5_22

[32] JUDITSKY, A. and NEMIROVSKI, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statist.* **28** 681–712. MR1792783 https://doi.org/10.1214/aos/1015951994

[33] KOMIYA, H. (1988). Elementary proof for Sion's minimax theorem. *Kodai Math. J.* **11** 5–7. MR0930413 https://doi.org/10.2996/kmj/1138038812

[34] KOOLEN, W. M., GRÜNWALD, P. D. and VAN ERVEN, T. (2016). Combining adversarial guarantees and stochastic fast rates in online learning. In *Advances in Neural Information Processing Systems* 29.

[35] KOOLEN, W. M. and VAN ERVEN, T. (2015). Second-order quantile methods for experts and combinatorial games. In *Proceedings of the* 28*th Conference on Learning Theory*.

[36] KOTŁOWSKI, W. (2018). On minimaxity of follow the leader strategy in the stochastic setting. *Theoret. Comput. Sci.* **742** 50–65. MR3844274 https://doi.org/10.1016/j.tcs.2017.12.030

[37] LAI, T. L. (1988). Nearly optimal sequential tests of composite hypotheses. *Ann. Statist.* **16** 856–886. MR0947582 https://doi.org/10.1214/aos/1176350840

[38] LATTIMORE, T. and SZEPESVÁRI, C. (2020). *Bandit Algorithms*. Cambridge Univ. Press, Cambridge.

[39] LITTLESTONE, N. and WARMUTH, M. K. (1994). The weighted majority algorithm. *Inform. and Comput.* **108** 212–261. MR1265851 https://doi.org/10.1006/inco.1994.1009

[40] LUO, H. and SCHAPIRE, R. E. (2015). Achieving all with no parameters: Adanormalhedge. In 28*th Annual Conference on Learning Theory*.

[41] LYKOURIS, T., MIRROKNI, V. and PAES LEME, R. (2018). Stochastic bandits robust to adversarial corruptions. In *STOC'*18*—Proceedings of the* 50*th Annual ACM SIGACT Symposium on Theory of Computing* 114–122. ACM, New York. MR3826238 https://doi.org/10.1145/3188745.3188918

[42] MCMAHAN, H. B. (2017). A survey of algorithms and analysis for adaptive online learning. *J. Mach. Learn. Res.* **18** Paper No. 90, 50 pp. MR3714253

[43] MCMAHAN, H. B. and STREETER, M. J. (2010). Adaptive bound optimization for online convex optimization. In *Proceedings of the* 23*rd Conference on Learning Theory*.

[44] MOURTADA, J. and GAÏFFAS, S. (2019). On the optimality of the hedge algorithm in the stochastic regime. *J. Mach. Learn. Res.* **20** Paper No. 83, 28 pp. MR3960937

[45] NEU, G. (2015). First-order regret bounds for combinatorial semi-bandits. In *Proceedings of the* 28*th Conference on Learning Theory*.

[46] RAKHLIN, A., SRIDHARAN, K. and TEWARI, A. (2011). Online learning: Stochastic, constrained, and smoothed adversaries. In *Advances in Neural Information Processing Systems* 25.

[47] RAKHLIN, A., SRIDHARAN, K. and TSYBAKOV, A. B. (2017). Empirical entropy, minimax regret and minimax risk. *Bernoulli* **23** 789–824. MR3606751 https://doi.org/10.3150/14-BEJ679

[48] RASMUSSEN, S. L. (1980). A Bayesian approach to a problem in sequential estimation. *Ann. Statist.* **8** 1229–1243. MR0594640

[49] ROBBINS, H. (1952). Some aspects of the sequential design of experiments. *Bull. Amer. Math. Soc.* **58** 527–535. MR0050246 https://doi.org/10.1090/S0002-9904-1952-09620-8

[50] SELDIN, Y. and SLIVKINS, A. (2014). One practical algorithm for both stochastic and adversarial bandits. In *Proceedings of the* 31*st International Conference on Machine Learning*.

[51] SION, M. (1958). On general minimax theorems. *Pacific J. Math.* **8** 171–176. MR0097026

[52] THUNE, T. and SELDIN, Y. (2018). Adaptation to easy data in prediction with limited advice. In *Advances in Neural Information Processing Systems* 31.

[53] TSYBAKOV, A. B. (2003). Optimal rates of aggregation. In *Proceedings of the* 16*th Conference on Computational Learning Theory*.

[54] TSYBAKOV, A. B. (2004). Optimal aggregation of classifiers in statistical learning. *Ann. Statist.* **32** 135–166. MR2051002 https://doi.org/10.1214/aos/1079120131

[55] VAN ERVEN, T., GRÜNWALD, P. D., KOOLEN, W. M. and DE ROOIJ, S. (2011). Adaptive hedge. In *Advances in Neural Information Processing Systems* 24.

[56] VAN ERVEN, T. and KOOLEN, W. M. (2016). Metagrad: Multiple learning rates in online learning. In *Advances in Neural Information Processing Systems* 29.

[57] VOVK, V. (1998). A game of prediction with expert advice. *J. Comput. System Sci.* **56** 153–173. MR1629690 https://doi.org/10.1006/jcss.1997.1556

[58] WALD, A. (1945). Sequential tests of statistical hypotheses. *Ann. Math. Stat.* **16** 117–186. MR0013275 https://doi.org/10.1214/aoms/1177731118

[59] WEI, C.-Y. and LUO, H. (2018). More adaptive algorithms for adversarial bandits. In *Proceedings of the* 31*st Annual Conference on Learning Theory*.

[60] WOLFOWITZ, J. (1947). Consistency of sequential binomial estimates. *Ann. Math. Stat.* **18** 131–135. MR0019899 https://doi.org/10.1214/aoms/1177730502

[61] YANG, Y. (2004). Aggregating regression procedures to improve performance. *Bernoulli* **10** 25–47. MR2044592 https://doi.org/10.3150/bj/1077544602

[62] ZIMMERT, J. and SELDIN, Y. (2019). An optimal algorithm for stochastic and adversarial bandits. In *The* 22*nd International Conference on Artificial Intelligence and Statistics*.