The Benefit of Being Bayesian in Online Conformal Prediction

Author Names Omitted for Anonymous Review. Paper-ID [add your ID here]

***Abstract*—Based on the framework of Conformal Prediction (CP), we study the online construction of confidence sets given a black-box machine learning model. By converting the target confidence levels into quantile levels, the problem can be reduced to predicting the quantiles (in hindsight) of a sequentially revealed data sequence. Two very different approaches have been studied previously:**

- **Assuming the data sequence is iid or exchangeable, one could maintain the empirical distribution of the observed data as an algorithmic belief, and directly predict its quantiles.**
- **Due to the fragility of statistical assumptions, a recent trend is to consider the adversarial setting and apply first-order online optimization to moving quantile losses. It requires the oracle knowledge of the target quantile level, and suffers from a previously overlooked monotonicity issue due to the associated loss linearization.**

This paper presents a CP algorithm that combines their strengths. Without any statistical assumption, it is able to answer multiple arbitrary confidence level queries with low regret, while also overcoming the monotonicity issue suffered by first-order optimization baselines. Furthermore, if the data sequence is indeed iid, then the same algorithm is automatically equipped with the "correct" coverage probability guarantee.

From a technical perspective, our key idea is to regularize the aforementioned algorithmic belief (the empirical distribution) by a Bayesian prior, which robustifies it by simulating a non-linearized Follow the Regularized Leader (FTRL) algorithm on the output. Such a belief update backbone is shared by prediction heads targeting different confidence levels, bringing practical benefits analogous to the recently proposed concept of U-calibration [17].

I. Introduction

Modern machine learning (ML) models are better at point prediction compared to probabilistic prediction. For example, when given an image classification task, they are better at responding "*this image is most likely a white cat*", rather than "*I'm 90% sure this image is an animal, 60% sure it's a cat, and 30% sure it's a white cat*". For downstream users, the more nuanced probabilistic predictions are often important for risk assessment. The challenge, however, lies in aligning the model's own uncertainty evaluation with its actual performance in the real world.

Conformal prediction (CP) [28] has recently emerged as a premier framework to address this challenge, as it blends the empirical strength of modern ML with the theoretical soundness of traditional statistical methods. As illustrated in Figure 1, CP algorithms make *confidence set predictions* on the label space, by sequentially interacting with three other parties: the *nature* (i.e., the data stream), a *black-box ML model*, and *downstream users*. Concretely, let $\mathcal{X}$ and $\mathcal{Y}$ be the covariate space and the label space. In each (the t-th) round, we study the following interaction protocol.

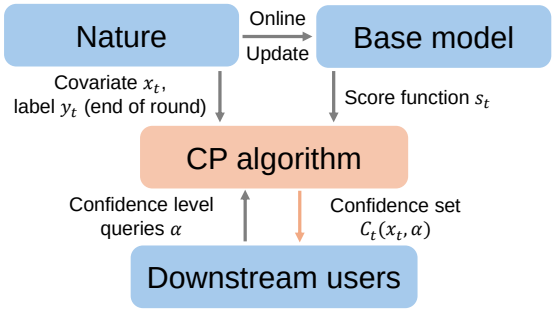

Fig. 1: An illustration of the CP interaction protocol.

1) We, as the CP algorithm, observe a *target covariate* $x_t \in \mathcal{X}$ from the nature, and a *score function* $s_t : \mathcal{X} \times \mathcal{Y} \to [0, R]$ generated by a black-box ML model called Base.
2) The downstream users select a finite set of *confidence level queries*, $A_t \subset [0, 1]$.
3) Given each confidence level query $\alpha \in A_t$, we predict a *score threshold* $r_t(x_t, \alpha)$ based on existing observations, which leads to a *confidence set prediction*[1]

$$\mathcal{C}_t(x_t, \alpha) = \{y \in \mathcal{Y} : s_t(x_t, y) \leq r_t(x_t, \alpha)\} . \quad (1)$$

4) Nature reveals the *ground truth label* $y_t \in \mathcal{Y}$, and thus the *true score* $r_t^* := s_t(x_t, y_t)$, to us.
5) The (x_t, y_t) pair is passed to Base, which it optionally uses to generate the score function s_{t+1}.

The intuition is that by sequentially evaluating Base on the target data, we generate better score thresholds that "correct" the uncertainty evaluation from Base itself. Our goal is thus clear in a very broad sense – predicting confidence sets with theoretically guaranteed *validity*. Say if a user only queries the confidence level $\alpha = 90\%$, then our CP algorithm needs to provide certain quantitative evidence that incentivizes the user to treat $\mathcal{C}_t(x_t, \alpha)$ as the 90% confidence set about the true label y_t. The same should hold true when the user queries $\alpha = 80\%, 95\%, 99\%, \ldots$, or all of these values at the same time. While this is well-studied under various statistical assumptions introduced later, the present work is mainly about the *adversarial*, game-theoretic setting where no statistical assumption is imposed at all.

[1]Without loss of generality, we assume the score function $s_t(x_t, y)$ is *negatively oriented* [24]: smaller $s_t(x_t, y)$ means the ML model is more certain that the candidate label y is the true label y_t.

A. Background

Before describing our main results, we need to present some necessary background. Let us start from the simplest case: assume that (i) within a time horizon T, the collection of true scores $r^*_{1:T}$ are iid samples of a random variable X with strictly positive density, and (ii) the α-*quantile* of X,[2] denoted by $q_\alpha(X)$, is known. A natural strategy is thus predicting $r_t(x_t, \alpha) = q_\alpha(X)$, which ensures that the *coverage condition* $y_t \in \mathcal{C}_t(x_t, \alpha)$ holds with probability exactly α. This certifies the validity of $\mathcal{C}_t(x_t, \alpha)$ in a strong probabilistic sense.

Although the imposed assumptions are clearly unrealistic, this example illustrates a central principle of CP: the predicted score threshold $r_t(x_t, \alpha)$ should ideally be the α-quantile of *some* distribution of $r^*_{1:T}$. A key challenge of CP is thus generalizing this principle to more realistic settings, as described below.

- *Direct approach:* Still assuming the sequence $r^*_{1:T}$ is iid but the population quantile $q_\alpha(X)$ is unknown, we could instead estimate $q_\alpha(X)$ on the fly. Specifically, our algorithm maintains the empirical distribution of $r^*_{1:t-1}$, denoted by $P_t = \bar{P}(r^*_{1:t-1})$, as an *algorithmic belief* about the unknown distribution of X. When queried with an arbitrary confidence level α, the algorithm "post-processes" the belief by setting $r_t(x_t, \alpha) = q_\alpha(P_t)$. In the standard terms of statistical learning, this is equivalent to *Empirical Risk Minimization* (ERM) with the *quantile loss* $l_\alpha(r, r^*) := (\mathbf{1}[r \geq r^*] - \alpha)(r - r^*)$, i.e.,
$$r_t(x_t, \alpha) = q_\alpha(P_t) \in \operatorname*{arg\,min}_{r \in [0,R]} \sum_{i=1}^{t-1} l_\alpha(r, r^*_i). \tag{2}$$
A well-known refinement of this approach called *Split Conformal* [23] sets $r_t(x_t, \alpha) = q_{\alpha + o(1)}(P_t)$. The $o(1)$ offset (wrt $t \to \infty$) ensures that even under a relaxation of the iid condition called *exchangeability*, a suitable notion of coverage probability is lower bounded by α [25].
- *Indirect approach:* Unfortunately, even the iid or exchangeability assumption often does not hold in practice.[3] Motivated by this challenge, a recent trend [10] is to remove all statistical assumptions, and instead estimate the empirical quantile of $r^*_{1:T}$ using first-order optimization algorithms from *adversarial online learning* [13, 22]. Taking gradient descent for example, such an approach amounts to picking an initialization $r_1(x_1, \alpha) \in [0, R]$ and following with the incremental update
$$r_{t+1}(x_{t+1}, \alpha) = r_t(x_t, \alpha) - \eta_t \partial l_\alpha(r_t(x_t, \alpha), r^*_t), \tag{3}$$
where $\eta_t > 0$ is the *learning rate*, and $\partial l_\alpha(r, r^*)$ can be any subgradient of the quantile loss l_α with respect to the first argument. Due to the absence of probability in this setting, alternative performance metrics have to be considered, such as the *post-hoc coverage frequency* and the *regret*.

[2]For the readers' reference, the α-quantile of a real random variable X is defined as $q_\alpha(X) := \min\{x : \mathbb{P}(X \leq x) \geq \alpha\}$.

[3]For example, when BASE itself is updated using gradient descent in Step 5 of the interaction protocol.

How do these two approaches compare? Although first-order optimization does not need statistical assumptions, it requires being "iterate-centric" rather than "data-centric": one needs to fix a single confidence level α beforehand, and the predicted threshold $r_t(x_t, \alpha)$ depends on how previous predictions $r_1(x_1, \alpha), \ldots, r_{t-1}(x_{t-1}, \alpha)$ compare to the true scores $r^*_{1:t-1}$, rather than just $r^*_{1:t-1}$ itself. Regarding the obtained confidence sets, this leads to a critical but previously overlooked *monotonicity issue:*

- Two copies of an algorithm with $\alpha_1 < \alpha_2$ can output $\mathcal{C}_t(x_t, \alpha_2) \subsetneq \mathcal{C}_t(x_t, \alpha_1)$, even if the initializations are the same. That is, the higher-confidence set is strictly smaller, violating the monotonicity requirement of probability measures. See Section II for details.

In contrast, the direct ERM approach does not suffer from this issue. The problem is that being equivalent to *Follow the Leader* (FTL) in online learning, it is well-known that ERM can suffer the vacuous $\Omega(T)$ regret on adversarial quantile losses. This motivates the question:

Is there an algorithm that enjoys the best of both worlds?

B. Our Result

Regarding the above question, we present a CP algorithm that combines several strengths of previous attempts:

- Just like the ERM approach, it can answer multiple arbitrary confidence level queries online.
- Without any statistical assumption, it guarantees the optimal regret bound
$$\sum_{t=1}^{T} l_\alpha(r_t(x_t, \alpha), r^*_t) - \sum_{t=1}^{T} l_\alpha(q_\alpha(r^*_{1:T}), r^*_t) = O(R\sqrt{T}),$$
simultaneously for all time horizon T, true score sequence $r^*_{1:T}$, and confidence level $\alpha \in [0, 1]$. Notice that the comparator $q_\alpha(r^*_{1:T})$ would be a natural fixed prediction had one known the empirical distribution of the true score sequence $r^*_{1:T}$ beforehand.
- Unlike first-order optimization baselines, it does not suffer from the aforementioned monotonicity issue, due to being "data-centric" rather than "iterate-centric".
- If the true scores $r^*_{1:T}$ are indeed iid, then it automatically achieves almost the same guarantees, including the *dataset-conditional coverage probability* and the *excess quantile risk*, as the ERM baseline. That is, the proposed algorithm can *adapt* to the easier statistical setting of CP, despite being designed for the more general adversarial setting.

From a technical perspective, our algorithm is a Bayesian modification of the ERM approach. Instead of setting its algorithmic belief as the empirical distribution of the past, $P_t = \bar{P}(r^*_{1:t-1})$, we set it as the convex combination
$$P_t = \lambda_t P_0 + (1 - \lambda_t)\bar{P}(r^*_{1:t-1}), \tag{4}$$
where P_0 is a prior, and $\lambda_t \in [0, 1]$ is a hyperparameter. The key observations are two-fold.

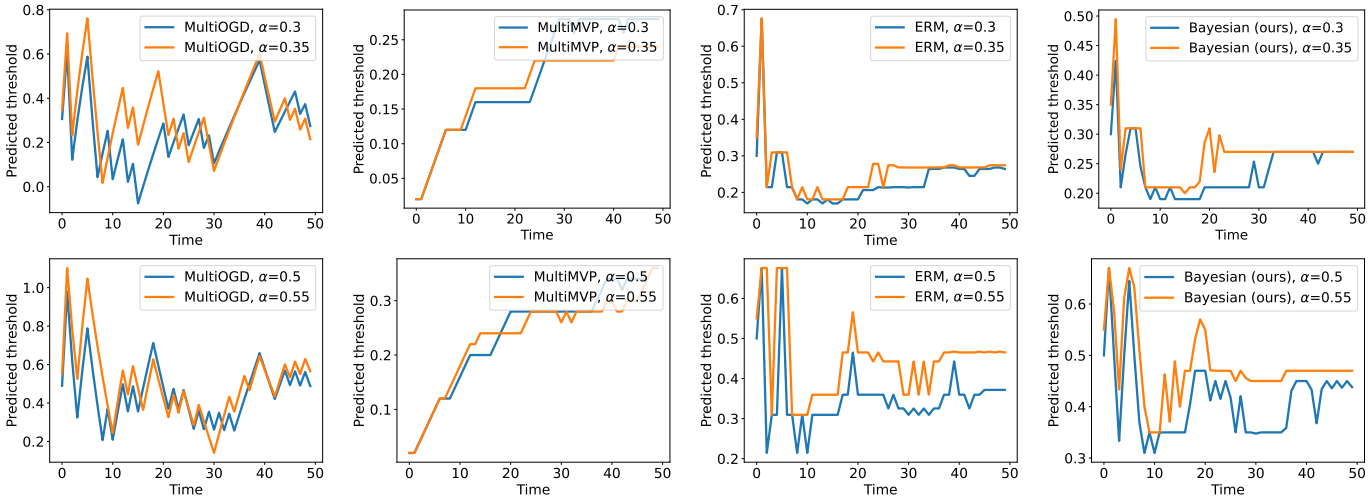

Fig. 2: Evaluating the monotonicity of threshold predictions. Ideally the orange line should be always above the blue line, since the associated target confidence level is higher. Columns correspond to different algorithms; rows correspond to different confidence levels.

- If $r^*_{1:t-1}$ are iid samples of an underlying distribution, then Eq.(4) is the posterior mean of a *Bayesian distribution estimator* based on the *Dirichlet process,* whereas the existing choice of $P_t = \bar{P}(r^*_{1:t-1})$ corresponds to the frequentist distribution estimator. Choosing $\lambda_t = O(t^{-1/2})$ leads to the state-of-the-art statistical guarantees.
- Without any statistical assumption, the Bayesian version of P_t offers the advantage of *downstream regularization*: the associated score threshold prediction $r_t(x_t, \alpha) = q_\alpha(P_t)$ is equivalent to the output of a non-linearized *Follow the Regularized Leader* (FTRL) algorithm from adversarial online learning. The desirable regret bound naturally follows, and the monotonicity issue is resolved as well.

Due to space constraints, a thorough discussion of related works is deferred to Appendix A.

C. Notation

This paper studies the *marginal* setting of CP, which means the threshold prediction $r_t(x_t, \alpha)$ will be independent of x_t; therefore we write it as $r_t(\alpha)$ for conciseness. For any symbol x, $x_{1:t}$ (e.g., $r^*_{1:t}$) represents the tuple $[x_1, \ldots, x_t]$. $\bar{P}(\cdot)$ denotes the empirical distribution of its input, and $q_\alpha(\cdot)$ denotes the α-quantile. Our regret bound concerns the quantile (or pinball) loss defined as $l_\alpha(r, r^*) := (\mathbf{1}[r \geq r^*] - \alpha)(r - r^*)$. log denotes the natural logarithm.

II. The Need for Monotonicity

To begin with, we use a numerical experiment to elaborate the monotonicity issue suffered by existing adversarial online CP algorithms. This has been overlooked in the literature, as all the existing approaches we are aware of require fixing a single target confidence level α at the beginning of the CP game.

Specifically, we consider two baselines, *Online Gradient Descent* (OGD) from [10], and *MultiValid Prediction* (MVP) from [3]. To make them support multiple confidence level queries, we adopt the following nearest-neighbor routing on top of their independent copies. It consists of three steps:

1) Evenly discretize the $[0, 1]$ interval of possible confidence levels using a grid $\tilde{A}$.
2) For each $\tilde{\alpha} \in \tilde{A}$, maintain a copy of the considered online CP algorithm (OGD or MVP) targeting $\tilde{\alpha}$.
3) When queried with α, follow the output of the "base" algorithm targeting the nearest neighbor of α in $\tilde{A}$.

Initiated with OGD and MVP, the resulting baselines are named as MultiOGD and MultiMVP respectively.

In the experiment, we fix $R = 1$. The true score sequence $r^*_{1:T}$ is sampled iid from the uniform distribution on $[0, 1]$, and we evaluate the thresholds $r_{1:T}(\alpha)$ predicted by different CP algorithms, under different α values. For each OGD baseline targeting $\tilde{\alpha}$, we use the standard learning rate $\eta_t = t^{-1/2}$, and initialize it with $r_1(\tilde{\alpha}) = \tilde{\alpha}$. The base MVP algorithms are all initialized at 0 following [3]. The point is that the initialization of MultiOGD and MultiMVP cannot cause any monotonicity violation. For comparison, we also test ERM as well as our Bayesian algorithm to be introduced in Section III.

The results are visualized in Figure 2. Ideally, in all the figures the orange line should be always above the blue line (i.e., the predicted confidence set due to Eq.(1) is larger), since the associated confidence level α is higher. Unlike ERM and our algorithm, both MultiOGD and MultiMVP violate this property, which hurts their trustworthiness to downstream users. We remark that although the data generation mechanism and the MultiMVP baseline are both randomized, a single random seed is used in this experiment to demonstrate the existence of the issue.

III. Algorithm and Analysis

Our main contribution is Algorithm 1, a Bayesian modification of the ERM approach. Let P_0 be an arbitrary distribution

Algorithm 1 Online CP with regularized belief.

Require: Step sizes $\{\lambda_t\}_{t\in\mathbb{N}_+}$, where $\lambda_1 = 1$, and $0 < \lambda_t < 1$ for all $t \geq 2$. A distribution P_0 on $[0 : R]$ called a prior, with strictly positive density function p_0.

for $t = 1, 2, \ldots$ **do**
  Compute the empirical distribution $\bar{P}(r^*_{1:t-1})$, and set the algorithmic belief P_t to

$$P_t = \lambda_t P_0 + (1 - \lambda_t)\bar{P}(r^*_{1:t-1}). \quad (5)$$

  for $\alpha \in A_t$ **do**
    Output the score threshold $r_t(\alpha) = q_\alpha(P_t)$.
  end for
  Observe the true score r^*_t.
end for

on $[0, R]$ called a *prior*, with a strictly positive density function p_0. In each round, Algorithm 1 maintains an algorithmic belief P_t by mixing P_0 with the empirical distribution of the previous true scores, $\bar{P}(r^*_{1:t-1})$. Then, given each queried confidence level α, it picks $r_t(\alpha) = q_\alpha(P_t)$ just like the ERM approach. We note that since P_t does not depend on any specific α, Algorithm 1 can naturally support multiple arbitrary confidence level queries online. By construction, it is also clear that for any $\alpha_1 < \alpha_2$ we always have $r_t(\alpha_1) \leq r_t(\alpha_2)$. That is, Algorithm 1 does not suffer from the aforementioned monotonicity issue.

To proceed, the first two subsections respectively analyze Algorithm 1 in the adversarial and the iid setting. They are followed by two extensions.

A. Analysis: Adversarial Setting

Without any statistical assumption, we view Algorithm 1 from the perspective of adversarial online learning. The key observation is that the mixture version of P_t induces *downstream regularizations* on the predicted threshold $r_t(\alpha)$, simultaneously for all $\alpha \in [0, 1]$. Furthermore, the downstream regularizers can take the surprisingly simple quadratic form due to the structure of the quantile loss.

Concretely, we define $\psi(r) := \mathbb{E}_{r^*\sim P_0}[l_\alpha(r, r^*)]$.

Theorem 1. *For all $\alpha \in [0, 1]$, the output $r_t(\alpha)$ of Algorithm 1 satisfies $r_1(\alpha) = \arg\min_{r\in\mathbb{R}} \psi(r)$, and $\forall t \geq 2$,*

$$r_t(\alpha) = \arg\min_{r\in\mathbb{R}} \left[\frac{\lambda_t(t-1)}{1-\lambda_t}\psi(r) + \sum_{i=1}^{t-1} l_\alpha(r, r^*_i) \right]. \quad (6)$$

Specifically,

- *ψ is strongly convex with coefficient $\inf_{r\in[0,R]} p_0(r)$, if the latter is positive.*
- *If P_0 is the uniform distribution on $[0, R]$, then ψ is a quadratic function centered at αR,*

$$\psi(r) = \frac{1}{2R}r^2 - \alpha r + \frac{1}{2}\alpha R.$$

Theorem 1 shows that despite not knowing α at the beginning of the CP game, Algorithm 1 generates the same output $r_t(\alpha)$ as a non-linearized *Follow the Regularized Leader* (FTRL) algorithm on the quantile loss l_α. Specifically, Eq.(6) can be compared to the FTL-equivalence of the iid-based approach, Eq.(2). The important difference is the additional regularizer $\psi(r)$.

To provide more context here: FTRL is a standard improvement of ERM / FTL in adversarial online learning, with better stability and worst-case performance on "difficult loss functions". Our analysis involves the non-linearized version of FTRL, which has previously received less attention than its linearized counterpart. This is largely due to computational reasons, since non-linearized FTRL has to solve a convex optimization subroutine in each round, whereas linearized FTRL admits closed-form solutions [22, Chapter 7.3]. From this perspective, a notable novelty of our result is showing that for a class of benign regularizers, non-linearized FTRL on quantile losses can be simulated by a simple and efficient procedure.

From Theorem 1, we can then obtain the regret bound of Algorithm 1 using the standard FTRL analysis. In order to demonstrate the role of good priors, the strong convexity of the regularizer ψ will be measured locally.

Theorem 2. *Let $\mu_{t,\alpha} := \inf\{p_0(r) : r_t(\alpha) \wedge r^*_t \leq r \leq r_t(\alpha) \vee r^*_t\}$. With the step size $\lambda_t = 1/\sqrt{t}$, Algorithm 1 guarantees*

$$\begin{aligned}\mathrm{Regret}_T(\alpha) &:= \sum_{t=1}^{T} l_\alpha(r_t(\alpha), r^*_t) - \sum_{t=1}^{T} l_\alpha(q_\alpha(r^*_{1:T}), r^*_t) \\ &= O\left(\psi(q_\alpha(r^*_{1:T}))\sqrt{T} + \sum_{t=1}^{T} \frac{1}{\mu_{t,\alpha}\sqrt{t}}\right), \quad (7)\end{aligned}$$

for all time horizon T, true score sequence $r^*_{1:T}$, and confidence level $\alpha \in [0, 1]$. Furthermore, if the prior P_0 is the uniform distribution on $[0, R]$, then

$$\mathrm{Regret}_T(\alpha) = O(R\sqrt{T}).$$

Let us interpret this regret bound. Suppose the time horizon T and the empirical true score distribution $\bar{P}(r^*_{1:T})$ are known beforehand (but the exact $r^*_{1:T}$ sequence is unknown), then for all α, a very natural strategy is to predict $r_t(\alpha) = q_\alpha(r^*_{1:T})$. Theorem 2 shows that without any statistical assumption, Algorithm 1 with the uniform prior asymptotically performs as well as this oracle in terms of the total quantile loss, and the $O(R\sqrt{T})$ regret bound is known to be tight [13, 22]. Existing first-order optimization baselines are equipped with regret bounds of a similar type [4, 11, 31], but the difference is that they require knowing the confidence level α beforehand, whereas Algorithm 1 achieves low regret simultaneously for all $\alpha \in [0, 1]$.

a) The role of good prior: A strength of Theorem 2 is that the $O(R\sqrt{T})$ regret bound only requires P_0 being the uniform distribution. Nonetheless, if one has extra prior knowledge on the environment, picking a more sophisticated prior can indeed bring advantages. To see this, notice that the function ψ in Eq.(7) is minimized at $q_\alpha(P_0)$, therefore ideally we would aim for $P_0 \approx \bar{P}(r^*_{1:T})$, which means $q_\alpha(P_0) \approx q_\alpha(r^*_{1:T})$ for all α. But unlike the discrete distribution $\bar{P}(r^*_{1:T})$, P_0 also needs to

have a "positive enough" density function, as otherwise the second term in Eq.(7) would blow up.

b) Coverage frequency error: We also note that existing first-order optimization baselines [4, 11, 31] are equipped with both a regret bound and a coverage frequency error bound, Eq.(8). Hoping to challenge this convention, here we discuss the advantages of only considering the regret.

First, the coverage frequency error is fundamentally "iterate-centric", whereas an ideal performance metric needs to be "data-centric". Consider the CP interaction protocol displayed in Figure 1: achieving low coverage frequency error requires the CP algorithm's output to depend not only on the top level (the nature and the base model), but also on the users' previous confidence level queries. This is in contrast with our regret minimization algorithm, whose output is independent of the users' query history.

Furthermore, just like the pathological example given by Bastani et al. [3], first-order optimization baselines essentially achieve the desirable coverage frequency due to the "overshooting" provided by the loss linearization. This is perhaps clear from the first online CP algorithm (ACI) proposed by Gibbs and Candès [10]: regarding the update Eq.(3) with the constant learning rate $\eta_t = \eta$, it is shown that the coverage frequency error monotonically decreases as $\eta \to \infty$. Such a peculiar behavior results precisely from overshooting: if $\alpha = 90\%$, then a failed coverage needs nine successful coverages to compensate, and *ensuring* this does not have much to do with the observed data. This casts some natural doubt on the coverage frequency error that the algorithm is designed to optimize.

c) Relation to U-calibration: Our results so far deviate from the common scope of online learning, which requires specifying a single loss function in each round. Instead, they have a similar flavor as *U-calibration* [17, 19]: forecasting for an unknown downstream agent. Prior works on U-calibration considered the setting of *finite-class distributional prediction* with generic proper losses, while our paper focuses on the continuous domain $[0, R]$ with the more specific quantile losses. The extra problem structure allows our algorithm to be deterministic (rather than being randomized like FTPL), thus establishing a closer connection to typical deterministic algorithms in *online convex optimization*.

B. Analysis: IID Setting

In practice, a CP algorithm is often applied without knowing the characteristics of the nature. We have been focusing on the adversarial setting, but what if the true scores $r^*_{1:T}$ turn out to be iid? We now demonstrate the *adaptivity* of Algorithm 1: it automatically achieves almost the same guarantees as ERM under the additional iid assumption.

First, as the coverage probability becomes the default performance metric in the iid setting, we present the following bound on the *dataset-conditional coverage probability*. Notice that the event of successful coverage can be expressed as $r^*_t \le r_t(\alpha)$, where $r_t(\alpha)$ is determined by the past true scores $r^*_{1:t-1}$ and the queried α.

Theorem 3. *Assume the true score sequence $r^*_1, r^*_2, \dots$ is drawn iid from an unknown continuous distribution $\mathcal{D}$. With the step size $\lambda_t = 1/\sqrt{t}$ and an arbitrary prior P_0, Algorithm 1 guarantees that for any fixed $t \ge 2$, with probability at least $1-\delta$ over the randomness of $r^*_{1:t-1}$, we have for all $\alpha \in [0,1]$,*

$$\alpha - \sqrt{\frac{\log(2/\delta)}{2(t-1)}} - \frac{1}{\sqrt{t}-1} \le \mathbb{P}_{r^*_t \sim \mathcal{D}}\left[r^*_t \le r_t(\alpha)\right] \le \alpha + \sqrt{\frac{\log(2/\delta)}{2(t-1)}} + \frac{1}{\sqrt{t}-1} + \frac{1}{t-1}.$$

Compared to the analogous result for ERM [25, Theorem 34], the difference here due to the regularization is the $(\sqrt{t}-1)^{-1}$ factor, which is dominated by the existing $O(\sqrt{t^{-1}\log\delta^{-1}})$ term resulting from the randomness. It shows that under the iid assumption, Algorithm 1 achieves almost the same coverage probability error as Split Conformal, despite being designed for the adversarial setting.

Besides the coverage probability, we can also analyze the *excess quantile risk* of Algorithm 1, which matches the standard oracle inequality one would obtain using ERM.

Theorem 4. *Assume the true score sequence $r^*_1, r^*_2, \dots$ is drawn iid from an unknown distribution $\mathcal{D}$. With the step size $\lambda_t = 1/\sqrt{t}$ and an arbitrary prior P_0, Algorithm 1 guarantees that for any fixed $t \ge 2$, with probability at least $1-\delta$ over the randomness of $r^*_{1:t-1}$, we have for all $\alpha \in [0,1]$,*

$$\mathbb{E}_{r^*_t \sim \mathcal{D}}[l_\alpha(r_t(\alpha), r^*_t)] \le \min_{r \in [0,R]} \mathbb{E}_{r^*_t \sim \mathcal{D}}[l_\alpha(r, r^*_t)] + O\left(R\sqrt{\frac{\log(1/\delta)}{t}}\right).$$

a) Bayesian interpretation: We have been calling Algorithm 1 "Bayesian". Following [9, Chapter 23], we now make this important connection concrete.

Consider the following distribution estimation problem: given $x_1, \dots, x_n \in \mathcal{X}$ sampled iid from an unknown distribution X, what is a good estimate of X? As opposed to the frequentist estimate $\bar{P}(x_{1:n})$, a Bayesian estimator would place a prior F_0 over all distributions supported on the domain $\mathcal{X}$, compute the posterior F_n from the samples, and output the mean $\mathbb{E}[F_n]$.

For analytical convenience, one would typically choose F_0 as a *conjugate prior*: it refers to a family of priors such that if F_0 belongs to this family, then F_n also belongs to this family. The most notable conjugate prior for distribution estimation is the *Dirichlet process* (DP), denoted as $\mathrm{DP}(\alpha, P_0)$. Here α and P_0 are hyperparameters: P_0 equals the mean $\mathbb{E}[\mathrm{DP}(\alpha, P_0)]$, while α controls the variance of $\mathrm{DP}(\alpha, P_0)$. Due to the conjugacy, if $F_0 = \mathrm{DP}(\alpha, P_0)$, then

$$F_n = \mathrm{DP}\left(\alpha + n, \frac{\alpha}{\alpha+n}P_0 + \frac{n}{\alpha+n}\bar{P}(x_{1:n})\right).$$

Consequently, the Bayesian estimator of the distribution X is

$$\mathbb{E}[F_n] = \frac{\alpha}{\alpha+n}P_0 + \frac{n}{\alpha+n}\bar{P}(x_{1:n}).$$

This is the same as the belief update Eq.(5) in Algorithm 1, with the hyperparameter $\lambda_t = \alpha/(\alpha+n)$. Our work can thus be regarded as the analysis of a Bayesian algorithm *with and without* the traditional iid assumption.

Back to the adversarial setting, we conclude this section with two extensions of Algorithm 1.

C. Extension: Quantization

Recall our construction of MultiOGD from Section II. Although not studied by existing works, it is not hard to see that with the size of the grid $\tilde{A}$ being $O(\sqrt{T})$, MultiOGD also satisfies the same α-agnostic $O(R\sqrt{T})$ regret bound as in Theorem 2, since the quantile loss $l_\alpha(r, r^*)$ is R-Lipschitz with respect to α. This raises a natural question: Algorithm 1 requires $O(T)$ memory due to storing the empirical distribution of previous true scores – can we reduce it to $O(\sqrt{T})$?

a) Quantized algorithm: Here is a variant of Algorithm 1, denoted as QUANTIZED, achieving this goal. The idea is to discretize the domain $[0, R]$ rather than the α-space: we maintain an evenly-spaced grid of size $\sqrt{T}$ over $[0, R]$, round each observed r_t^* to its nearest neighbor $\tilde{r}_t^*$ on the grid, and replace the belief update Eq.(5) by

$$P_t = \lambda_t P_0 + (1-\lambda_t)\bar{P}(\tilde{r}^*_{1:t-1}).$$

The associated regret bound follows from the Lipschitzness of $l_\alpha(r, r^*)$ with respect to r^*.

Theorem 5. *With $\lambda_t = 1/\sqrt{t}$ and the uniform P_0,* QUANTIZED *achieves* $\mathrm{Regret}_T(\alpha) = O(R\sqrt{T})$.

Compared to MultiOGD, QUANTIZED achieves the same $O(R\sqrt{T})$ regret bound with $O(\sqrt{T})$ memory, while avoiding its monotonicity issue. There is another practical advantage: after observing each r_t^*, MultiOGD needs to update all $\sqrt{T}$ base algorithms, whereas QUANTIZED performs only one update on the algorithmic belief $\bar{P}_t$, and then makes $|A_t|$ inferences using the prediction head.

D. Extension: Continual Distribution Shift

Starting from [10], the study of adversarial online CP has been largely motivated by the prevalence of continual distribution shifts in practice. Tackling this challenge requires *non-converging* algorithms characterized by sliding-window performance guarantees. We now present a discounted variant of Algorithm 1, denoted by DISCOUNTED, along this direction.

a) Discounted algorithm: Let $\beta \in (0,1)$ be a *discount factor*, which is a bandwidth hyperparameter required by DISCOUNTED. Then, we define a regularized and discounted empirical distribution of $r^*_{1:t}$ recursively by

$$\bar{P}_\beta(r_1^*) = \beta P_0 + (1-\beta)\delta(r_1^*),$$

$$\begin{aligned}\bar{P}_\beta(r^*_{1:t}) &= \beta \bar{P}_\beta(r^*_{1:t-1}) + (1-\beta)\delta(r_t^*) \\ &= \beta^t P_0 + (1-\beta)\sum_{i=1}^{t}\beta^{t-i}\delta(r_i^*),\end{aligned}$$

where $\delta(r_t^*)$ is the distribution with point mass at r_t^*. This is used to replace the undiscounted empirical distribution in the belief update, i.e., Eq.(5) is replaced by

$$P_t = \lambda_t P_0 + (1-\lambda_t)\bar{P}_\beta(r^*_{1:t-1}).$$

After that, the prediction head remains unchanged, i.e., $r_t(\alpha) = q_\alpha(P_t)$.

Similar to Theorem 1 and 2, we can show that DISCOUNTED simulates the β-discounted non-linearized FTRL, which is equipped with a β-discounted regret bound. Importantly, reasonable step sizes λ_t become constant (rather than decreasing), which emphasizes the crucial role of the prior P_0: instead of only using P_0 to regularize the beginning of the CP game, DISCOUNTED continually mix P_0 into its algorithm belief with constant weight, such that it does not "overfit the current environment".

Theorem 6. *With $\lambda_t = \lambda = \frac{\sqrt{1-\beta}}{\beta+\sqrt{1-\beta}}$ and the uniform P_0, the output $r_t(\alpha)$ of* DISCOUNTED *satisfies*

$$r_t(\alpha) = \arg\min_{r\in\mathbb{R}}\left[(1-\beta)^{-1}\left(\frac{\lambda}{1-\lambda}+\beta^{t-1}\right)\psi(r) + \sum_{i=1}^{t-1}\beta^{t-1-i} l_\alpha(r, r_i^*)\right],$$

for all α and t. In addition, for all $\alpha \in [0,1]$, it guarantees the discounted regret bound

$$\begin{aligned}&\mathrm{Regret}_{T,\beta}(\alpha) \\ &:= \sum_{t=1}^{T}\beta^{T-t} l_\alpha(r_t(\alpha), r_t^*) - \min_{r\in[0,R]}\sum_{t=1}^{T}\beta^{T-t} l_\alpha(r, r_t^*) \\ &\qquad \le \frac{R}{\sqrt{1-\beta}} + o(R),\end{aligned}$$

where $o(\cdot)$ is with respect to $T \to \infty$.

We remark that [31, Theorem 7] presents a discounted regret lower bound on linear losses, which can be converted to $\Omega(\min\{\alpha, 1-\alpha\}R/\sqrt{1-\beta^2})$ on the quantile losses we consider. Since $(1-\beta)^{-1/2} \le 2(1-\beta^2)^{-1/2}$ for all $\beta \in (0,1)$, Theorem 6 matches this lower bound in the minimax sense (with respect to α, i.e., when $\alpha = 1/2$).

b) Experiment: Finally, we conducted more experiments to validate the empirical performance of the proposed algorithms. Results are presented in Appendix D.

IV. CONCLUSION

Focusing on the online adversarial formulation of conformal prediction, we propose a novel Bayesian algorithm with a number of strengths: it supports multiple arbitrary confidence level queries, achieves low regret, avoids the monotonicity issue on the obtained confidence sets, and adapts to iid environments. We further develop quantized and discounted extensions of this algorithm, and our theoretical results are supported by experiments.

References

[1] Anastasios N Angelopoulos and Stephen Bates. Conformal prediction: A gentle introduction. *Foundations and Trends® in Machine Learning*, 16(4):494–591, 2023.

[2] Anastasios N Angelopoulos, Rina Foygel Barber, and Stephen Bates. Online conformal prediction with decaying step sizes. In *International Conference on Machine Learning*. PMLR, 2024.

[3] Osbert Bastani, Varun Gupta, Christopher Jung, Georgy Noarov, Ramya Ramalingam, and Aaron Roth. Practical adversarial multivalid conformal prediction. *Advances in Neural Information Processing Systems*, 35:29362–29373, 2022.

[4] Aadyot Bhatnagar, Huan Wang, Caiming Xiong, and Yu Bai. Improved online conformal prediction via strongly adaptive online learning. In *International Conference on Machine Learning*, pages 2337–2363. PMLR, 2023.

[5] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

[6] Aryeh Dvoretzky, Jack Kiefer, and Jacob Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, pages 642–669, 1956.

[7] Edwin Fong and Chris C Holmes. Conformal bayesian computation. *Advances in Neural Information Processing Systems*, 34:18268–18279, 2021.

[8] Sumegha Garg, Christopher Jung, Omer Reingold, and Aaron Roth. Oracle efficient online multicalibration and omniprediction. In *Symposium on Discrete Algorithms*, pages 2725–2792. SIAM, 2024.

[9] Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. Bayesian data analysis. http://www.stat.columbia.edu/~gelman/book/BDA3.pdf, 2021.

[10] Isaac Gibbs and Emmanuel Candès. Adaptive conformal inference under distribution shift. *Advances in Neural Information Processing Systems*, 34:1660–1672, 2021.

[11] Isaac Gibbs and Emmanuel J Candès. Conformal inference for online prediction with arbitrary distribution shifts. *Journal of Machine Learning Research*, 25(162): 1–36, 2024.

[12] Parikshit Gopalan, Adam Tauman Kalai, Omer Reingold, Vatsal Sharan, and Udi Wieder. Omnipredictors. In *Innovations in Theoretical Computer Science*, 2022.

[13] Elad Hazan. Introduction to online convex optimization. *arXiv preprint arXiv:1909.05207v3*, 2023.

[14] Ursula Hébert-Johnson, Michael Kim, Omer Reingold, and Guy Rothblum. Multicalibration: Calibration for the (computationally-identifiable) masses. In *International Conference on Machine Learning*, pages 1939–1948. PMLR, 2018.

[15] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[16] Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

[17] Bobby Kleinberg, Renato Paes Leme, Jon Schneider, and Yifeng Teng. U-calibration: Forecasting for an unknown agent. In *Conference on Learning Theory*, pages 5143–5145. PMLR, 2023.

[18] Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[19] Haipeng Luo, Spandan Senapati, and Vatsal Sharan. Optimal multiclass U-calibration error and beyond. In *Advances in Neural Information Processing Systems*, 2024.

[20] Pascal Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The annals of Probability*, pages 1269–1283, 1990.

[21] Radford M Neal. *Bayesian learning for neural networks*. Springer, 2012.

[22] Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2023.

[23] Harris Papadopoulos, Kostas Proedrou, Volodya Vovk, and Alex Gammerman. Inductive confidence machines for regression. In *European Conference on Machine Learning*, pages 345–356. Springer, 2002.

[24] Yaniv Romano, Matteo Sesia, and Emmanuel Candes. Classification with valid and adaptive coverage. *Advances in Neural Information Processing Systems*, 33:3581–3591, 2020.

[25] Aaron Roth. Uncertain: Modern topics in uncertainty estimation. https://www.cis.upenn.edu/~aaroth/uncertainty-notes.pdf, 2022.

[26] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3-4):285–294, 1933.

[27] Ryan Tibshirani. Advanced topics in statistical learning: Conformal prediction. https://www.stat.berkeley.edu/~ryantibs/statlearn-s23/lectures/conformal.pdf, 2023.

[28] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Algorithmic learning in a random world*, volume 29. Springer, 2005.

[29] Yunbei Xu and Assaf Zeevi. Bayesian design principles for frequentist sequential learning. In *International Conference on Machine Learning*, pages 38768–38800. PMLR, 2023.

[30] Tong Zhang. *Mathematical analysis of machine learning algorithms*. Cambridge University Press, 2023.

[31] Zhiyu Zhang, David Bombara, and Heng Yang. Discounted adaptive online learning: Towards better regularization. In *International Conference on Machine Learning*, pages 58631–58661. PMLR, 2024.

APPENDIX

A. Related Work

For the background of CP and its numerous applications, the readers are referred to several excellent resources [1, 25, 27, 28].

a) Online CP: Our work contributes to the growing body of literature on online adversarial conformal prediction, initiated by [10]. They showed that running gradient descent with constant learning rate can achieve low *coverage frequency error*, i.e.,

$$\left|\alpha - T^{-1}\sum_{t=1}^{T}\mathbf{1}[r_t^* \le r_t(x_t,\alpha)]\right| = o(1), \tag{8}$$

and its sliding-window analogues. [3] demonstrated an intriguing weakness of this performance metric: one could trivially satisfy this coverage frequency bound by predicting a data-independent alternation between the empty set and the entire label space. To rule out such pathological cases, it has thus been standard to consider an additional performance metric, such as the regret [4, 11, 31] or the *multi-calibrated* coverage frequency [3]. Under the additional iid assumption, [2] studied the asymptotic coverage probability achieved by gradient descent; in contrast, our best-of-both-world guarantees are all non-asymptotic.

Different from the literature, the present work emphasizes regret minimization in online CP, as we believe it offers advantages *even over simultaneously bounding the regret and the coverage frequency error* (since loss linearization is not necessary anymore). See Section III-A for a detailed discussion.

b) Bayesian uncertainty quantification: Our work is also drastically different from the typical uncertainty quantification pipeline in *Bayesian machine learning* [21], where one directly constructs posteriors on the parameter space and then uses it for confidence set predictions. While such a pipeline often suffers from scalability issues, we embed the construction of posteriors into the conformal prediction framework, thereby making the resulting algorithm both modular (i.e., can operate on top of arbitrary black-box ML models) and computationally efficient. Related but different from our focus, [7] studied a CP problem where the base ML model itself is Bayesian.

c) Adversarial Bayes: A common challenge for Bayesian algorithms is the choice of the prior. In this work, we exploit the algorithmic equivalence of Bayesian priors and regularizers, thereby utilizing the systematic study of regularization in adversarial online learning to demonstrate that *the uniform prior is minimax optimal* for regret minimization. That is, one could avoid tuning the prior in a theoretically justified manner. Broadly speaking, such an idea of "adversarial Bayes" is closely related to a classical algorithm family in online learning called *Follow the Perturbed Leader* (FTPL) [16]. Notable examples of FTPL include *Thompson sampling* [18, 26, 29], a prevalent Bayesian approach for bandits and reinforcement learning, and *U-calibration* [17, 19], a recently proposed framework for loss-agnostic decision making. Despite being deterministic, our approach resembles the high level idea of U-calibration. See Section III-A for a detailed discussion of their connections and differences.

Besides U-calibration, *omniprediction* [8, 12] is another iconic framework for loss-agnostic decision making, whose main idea is to maintain a *multi-calibrated* algorithmic belief in the sense of [14]. Our approach does not require calibration as an underlying mechanism.

B. Omitted Proofs

Theorem 1. *For all $\alpha \in [0,1]$, the output $r_t(\alpha)$ of Algorithm 1 satisfies $r_1(\alpha) = \arg\min_{r\in\mathbb{R}} \psi(r)$, and $\forall t \ge 2$,*

$$r_t(\alpha) = \arg\min_{r\in\mathbb{R}}\left[\frac{\lambda_t(t-1)}{1-\lambda_t}\psi(r) + \sum_{i=1}^{t-1} l_\alpha(r, r_i^*)\right]. \tag{6}$$

Specifically,

- *ψ is strongly convex with coefficient $\inf_{r\in[0,R]} p_0(r)$, if the latter is positive.*
- *If P_0 is the uniform distribution on $[0,R]$, then ψ is a quadratic function centered at αR,*

$$\psi(r) = \frac{1}{2R}r^2 - \alpha r + \frac{1}{2}\alpha R.$$

Proof of Theorem 1: We first rewrite the base regularizer ψ as

$$\begin{aligned}\psi(r) &= \int_0^R l_\alpha(r, r^*)p_0(r^*)dr^* \\ &= (1-\alpha)\int_0^r (r - r^*)p_0(r^*)dr^* + \alpha\int_r^R (r^* - r)p_0(r^*)dr^*.\end{aligned}$$

It is twice-differentiable, with

$$\psi'(r) = (1-\alpha)\int_0^r p_0(r^*)dr^* - \alpha\int_r^R p_0(r^*)dr^* = \int_0^r p_0(r^*)dr^* - \alpha,$$

and $\psi''(r) = p_0(r)$. The strong convexity statement on ψ is thus clear. If P_0 is uniform, we have

$$\begin{aligned}\psi(r) &= R^{-1}\left[(1-\alpha)\int_0^r (r-r^*)dr^* + \alpha\int_r^R (r^*-r)dr^*\right] \\ &= \frac{1}{2R}\left[(1-\alpha)r^2 + \alpha(R-r)^2\right] = \frac{1}{2R}r^2 - \alpha r + \frac{1}{2}\alpha R.\end{aligned}$$

Next, consider the first part of the theorem. The case of $t = 1$ is straightforward to verify. For any $t \geq 2$, Algorithm 1 outputs

$$\begin{aligned}r_t(\alpha) &= q_\alpha\left[\lambda_t P_0 + (1-\lambda_t)\bar{P}(r^*_{1:t-1})\right] \\ &= \min\left\{r : \lambda_t \int_0^r p_0(r^*)dr^* + \frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \leq r] \geq \alpha\right\}.\end{aligned} \tag{9}$$

On the other hand, consider the optimization objective in Eq.(6), which we write as

$$F_t(r) := \frac{\lambda_t(t-1)}{1-\lambda_t}\psi(r) + \sum_{i=1}^{t-1} l_\alpha(r, r_i^*). \tag{10}$$

Notice that the function $F_t(r)$ is continuous and right-differentiable. Taking its right-derivative, we have

$$\begin{aligned}F'_{t,+}(r) &= \frac{\lambda_t(t-1)}{1-\lambda_t}\left[\int_0^r p_0(r^*)dr^* - \alpha\right] + \left(-\alpha\sum_{i=1}^{t-1}\mathbf{1}[r < r_i^*] + (1-\alpha)\sum_{i=1}^{t-1}\mathbf{1}[r \geq r_i^*]\right) \\ &= \frac{\lambda_t(t-1)}{1-\lambda_t}\int_0^r p_0(r^*)dr^* - \frac{\alpha\lambda_t(t-1)}{1-\lambda_t} - \alpha(t-1) + \sum_{i=1}^{t-1}\mathbf{1}[r \geq r_i^*] \\ &= \frac{t-1}{1-\lambda_t}\left(\lambda_t\int_0^r p_0(r^*)dr^* + \frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r \geq r_i^*] - \alpha\right).\end{aligned}$$

Comparing it to Eq.(9), we see that the output $r_t(\alpha)$ of Algorithm 1, given by Eq.(9), satisfies

$$r_t(\alpha) = \min\{r : F'_{t,+}(r) \geq 0\}.$$

Since the function $F_t(r)$ is strictly convex, we have $r_t(\alpha) = \arg\min_r F_t(r)$, which is equivalent to Eq.(6). ■

Theorem 2. *Let $\mu_{t,\alpha} := \inf\{p_0(r) : r_t(\alpha) \wedge r_t^* \leq r \leq r_t(\alpha) \vee r_t^*\}$. With the step size $\lambda_t = 1/\sqrt{t}$, Algorithm 1 guarantees*

$$\begin{aligned}\mathrm{Regret}_T(\alpha) &:= \sum_{t=1}^T l_\alpha(r_t(\alpha), r_t^*) - \sum_{t=1}^T l_\alpha(q_\alpha(r^*_{1:T}), r_t^*) \\ &= O\left(\psi(q_\alpha(r^*_{1:T}))\sqrt{T} + \sum_{t=1}^T \frac{1}{\mu_{t,\alpha}\sqrt{t}}\right),\end{aligned} \tag{7}$$

for all time horizon T, true score sequence $r^*_{1:T}$, and confidence level $\alpha \in [0,1]$. Furthermore, if the prior P_0 is the uniform distribution on $[0, R]$, then

$$\mathrm{Regret}_T(\alpha) = O(R\sqrt{T}).$$

Proof of Theorem 2: The proof can be decomposed into the following steps.

a) Step 1: Starting from the FTRL formulation Eq.(6), we first verify that the regularizer weight $\frac{\lambda_t(t-1)}{1-\lambda_t}$ is increasing with respect to t (when $t > 1$), which is required by the FTRL analysis. To this end, define

$$h_t := \frac{\lambda_t(t-1)}{1-\lambda_t} = \frac{t-1}{\sqrt{t}-1}.$$

Taking the derivative with respect to t, for all $t > 1$,

$$\frac{dh_t}{dt} = \frac{\sqrt{t} - 1 - \frac{t-1}{2\sqrt{t}}}{(\sqrt{t}-1)^2} = \frac{t - 2\sqrt{t} + 1}{2\sqrt{t}(\sqrt{t+1}-1)^2} = \frac{(\sqrt{t}-1)^2}{2\sqrt{t}(\sqrt{t+1}-1)^2} \geq 0.$$

For completeness, we also define $h_1 = 1$.

Besides, we have the order estimate $h_t = O(\sqrt{t})$, $1/h_t = O(1/\sqrt{t})$, where $O(\cdot)$ only hides an absolute constant.

b) Step 2: Next, due to Theorem 1, we can apply the standard FTRL analysis. Recall our notation from Eq.(10): we write the optimization objective in Eq.(6) as

$$F_t(r) := h_t\psi(r) + \sum_{i=1}^{t-1} l_\alpha(r, r_i^*), \quad \forall t \geq 2.$$

Similarly, we also write $F_1(r) := h_1\psi(r)$. Notice that $r_t(\alpha) = \arg\min_{r\in\mathbb{R}} F_t(r)$ for all t.

The classical FTRL equality lemma [22, Lemma 7.1] states that

$$\begin{aligned}\mathrm{Regret}_T(\alpha) &= h_{T+1}\psi(q_\alpha(r_{1:T}^*)) - \min_{r\in\mathbb{R}}\psi(r) + \sum_{t=1}^{T}\left[F_t(r_t(\alpha)) - F_{t+1}(r_{t+1}(\alpha)) + l_\alpha(r_t(\alpha), r_t^*)\right] \\ &\quad + F_{T+1}(r_{T+1}(\alpha)) - F_{T+1}(q_\alpha(r_{1:T}^*)) \\ &\leq h_{T+1}\psi(q_\alpha(r_{1:T}^*)) + \sum_{t=1}^{T}\left[F_t(r_t(\alpha)) - F_{t+1}(r_{t+1}(\alpha)) + l_\alpha(r_t(\alpha), r_t^*)\right],\end{aligned}$$

where the second line is due to $\min_r \psi(r) \geq 0$, and $r_{T+1}(\alpha) = \arg\min_{r\in\mathbb{R}} F_{T+1}(r)$.

Consider the sum on the RHS, where for conciseness we omit (α) in the notation. This is the typical one-step quantity involved in the FTRL analysis. Following a similar procedure as [22, Lemma 7.8], we have

$$\begin{aligned}&F_t(r_t) - F_{t+1}(r_{t+1}) + l_\alpha(r_t, r_t^*) \\ &= F_t(r_t) + l_\alpha(r_t, r_t^*) - F_t(r_{t+1}) - l_\alpha(r_{t+1}, r_t^*) + (h_t - h_{t+1})\psi(r_{t+1}) \\ &\leq F_t(r_t) + l_\alpha(r_t, r_t^*) - F_t(r_{t+1}) - l_\alpha(r_{t+1}, r_t^*) \qquad (h_{t+1} \geq h_t,\ \psi(r_{t+1}) \geq 0) \\ &\leq F_t(r_t) + l_\alpha(r_t, r_t^*) - \min_{r\in\mathbb{R}}\left[F_t(r) + l_\alpha(r, r_t^*)\right].\end{aligned}$$

Observe that since $F_t(\cdot)$ and $l_\alpha(\cdot, r_t^*)$ are both convex, the minimizing argument of their sum lies between their respective unique minimizers, r_t and r_t^*. On this segment, the function F_t is $h_t\mu_{t,\alpha}$-strongly-convex, where $\mu_{t,\alpha}$ is defined in the assumption of the theorem. We now proceed using the property of strong convexity [22, Lemma 7.6], which we restate as Lemma A.1.

Concretely, if g_t is a subgradient of $l_\alpha(\cdot, r_t^*)$ at r_t, then it is also a subgradient of $F_t(\cdot) + l_\alpha(\cdot, r_t^*)$ at r_t, since $r_t = \arg\min_r F_t(r)$. Moreover, such a subgradient g_t satisfies $|g_t| \leq 1$ due to $l_\alpha(\cdot, r_t^*)$ being 1-Lipschitz. Combining these with the strong convexity, Lemma A.1 yields

$$F_t(r_t) + l_\alpha(r_t, r_t^*) - \min_{r\in\mathbb{R}}\left[F_t(r) + l_\alpha(r, r_t^*)\right] \leq \frac{1}{2h_t\mu_{t,\alpha}}.$$

Plugging this all the way back into the regret bound, we have

$$\begin{aligned}\mathrm{Regret}_T(\alpha) &\leq h_{T+1}\psi(q_\alpha(r_{1:T}^*)) + \frac{1}{2}\sum_{t=1}^{T}\frac{1}{h_t\mu_{t,\alpha}} \\ &= O\left(\psi(q_\alpha(r_{1:T}^*))\sqrt{T} + \sum_{t=1}^{T}\frac{1}{\mu_{t,\alpha}\sqrt{t}}\right).\end{aligned}$$

c) Step 3: Finally we analyze the special case of uniform prior. From Theorem 1,

$$\psi(q_\alpha(r_{1:T}^*)) \leq \max_{r\in[0,R]}\psi(r) \leq \max_{r\in[0,R]}\left(\frac{1}{2R}r^2 - \alpha r + \frac{1}{2}\alpha R\right) \leq \frac{R}{2}.$$

Furthermore, $\mu_{t,\alpha} = 1/R$. Plugging in $\sum_{t=1}^{T} t^{-1/2} = O(\sqrt{T})$ completes the proof. ■

Theorem 3. *Assume the true score sequence $r_1^*, r_2^*, \ldots$ is drawn iid from an unknown continuous distribution $\mathcal{D}$. With the step size $\lambda_t = 1/\sqrt{t}$ and an arbitrary prior P_0, Algorithm 1 guarantees that for any fixed $t \geq 2$, with probability at least $1-\delta$ over the randomness of $r_{1:t-1}^*$, we have for all $\alpha \in [0,1]$,*

$$\alpha - \sqrt{\frac{\log(2/\delta)}{2(t-1)}} - \frac{1}{\sqrt{t}-1} \leq \mathbb{P}_{r_t^*\sim\mathcal{D}}\left[r_t^* \leq r_t(\alpha)\right] \leq \alpha + \sqrt{\frac{\log(2/\delta)}{2(t-1)}} + \frac{1}{\sqrt{t}-1} + \frac{1}{t-1}.$$

Proof of Theorem 3: The proof follows a similar strategy as [25, Theorem 34]. First, for any fixed $t \geq 2$, the samples $r_{1:t-1}^*$ have no ties almost surely, since the underlying distribution $\mathcal{D}$ is continuous. We will condition the rest of the analysis on this event.

Next, recall Algorithm 1's prediction rule, Eq.(9). On one hand, we have

$$\lambda_t \int_0^{r_t(\alpha)} p_0(r^*)dr^* + \frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)] \ge \alpha,$$

which means

$$\frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)] \ge \alpha - \lambda_t,$$

$$\frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)] \ge \alpha + \frac{\lambda_t}{1-\lambda_t}(\alpha-1) \ge \alpha - \frac{1}{\sqrt{t}-1}.$$

On the other hand, if we define $m = \sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)]$ and let r_{-1}^* be the $(m-1)$-th smallest element of $r_{1:t-1}^*$, then it is also clear from Eq.(9) that

$$\lambda_t \int_0^{r_{-1}^*} p_0(r^*)dr^* + \frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_{-1}^*] \le \alpha,$$

which means

$$\frac{1-\lambda_t}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_{-1}^*] \le \alpha - \lambda_t \int_0^{r_{-1}^*} p_0(r^*)dr^* \le \alpha,$$

$$\frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_{-1}^*] \le \frac{\alpha}{1-\lambda_t} \le \alpha + \frac{1}{\sqrt{t}-1},$$

$$\frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)] \le \frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_{-1}^*] + \frac{1}{t-1} \le \alpha + \frac{1}{\sqrt{t}-1} + \frac{1}{t-1}.$$

In summary,

$$\alpha - \frac{1}{\sqrt{t}-1} \le \frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)] \le \alpha + \frac{1}{\sqrt{t}-1} + \frac{1}{t-1}. \tag{11}$$

Finally we apply the DKW inequality (Lemma A.3). For all $\varepsilon > 0$, we have

$$\mathbb{P}_{r_{1:t-1}^*}\left[\sup_{\alpha\in[0,1]}\left|\left(\frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)]\right) - \mathbb{P}_{r_t^*}[r_t^* \le r_t(\alpha)]\right| > \varepsilon\right] \le 2\exp\left[-2(t-1)\varepsilon^2\right].$$

Therefore, with probability at least $1-\delta$ over the randomness of $r_{1:t-1}^*$, we have

$$\left|\left(\frac{1}{t-1}\sum_{i=1}^{t-1}\mathbf{1}[r_i^* \le r_t(\alpha)]\right) - \mathbb{P}_{r_t^*}[r_t^* \le r_t(\alpha)]\right| \le \sqrt{\frac{\log(2/\delta)}{2(t-1)}}, \quad \forall \alpha \in [0,1].$$

Combining it with Eq.(11) above completes the proof. ■

Theorem 4. *Assume the true score sequence $r_1^*, r_2^*, \ldots$ is drawn iid from an unknown distribution $\mathcal{D}$. With the step size $\lambda_t = 1/\sqrt{t}$ and an arbitrary prior P_0, Algorithm 1 guarantees that for any fixed $t \ge 2$, with probability at least $1-\delta$ over the randomness of $r_{1:t-1}^*$, we have for all $\alpha \in [0,1]$,*

$$\mathbb{E}_{r_t^*\sim\mathcal{D}}[l_\alpha(r_t(\alpha), r_t^*)] \le \min_{r\in[0,R]} \mathbb{E}_{r_t^*\sim\mathcal{D}}[l_\alpha(r, r_t^*)] + O\left(R\sqrt{\frac{\log(1/\delta)}{t}}\right).$$

Proof of Theorem 4: The proof follows from a standard uniform convergence argument [30] combined with the Lipschitzness of the quantile loss.

First, notice that with any combination of α, r and r^*, the quantile loss $l_\alpha(r, r^*) \in [0, R]$. Therefore, fixing any $\alpha \in [0,1]$ and $r \in [0, R]$, we apply the Hoeffding's inequality (Lemma A.2) to obtain

$$\mathbb{P}_{r_{1:t-1}^*}\left[\left|\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r, r_i^*) - \mathbb{E}_{r_t^*}[l_\alpha(r, r_t^*)]\right| \ge \varepsilon\right] \le 2\exp\left(-\frac{2(t-1)\varepsilon^2}{R^2}\right).$$

Next, we evenly discretize $[0,1]$ by a grid of size $\sqrt{t}$, and also $[0,R]$ by a grid of size $\sqrt{t}$, and denote their combination as a set S. $|S|=t$. For all α and r, there exists $(\tilde{\alpha},\tilde{r})\in S$ satisfying $|\alpha-\tilde{\alpha}|\le 1/\sqrt{t}$ and $|r-\tilde{r}|\le R/\sqrt{t}$. Applying the union bound on S yields

$$\mathbb{P}_{r^*_{1:t-1}}\left[\max_{(\alpha,r)\in S}\left|\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)-\mathbb{E}_{r_t^*}[l_\alpha(r,r_t^*)]\right|\ge\varepsilon\right]\le 2t\exp\left(-\frac{2(t-1)\varepsilon^2}{R^2}\right),$$

which means with probability at least $1-\delta$,

$$\max_{(\alpha,r)\in S}\left|\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)-\mathbb{E}_{r_t^*}[l_\alpha(r,r_t^*)]\right|\le R\sqrt{\frac{\log(2t/\delta)}{2(t-1)}}.$$

Since $l_\alpha(r,r^*)$ is R-Lipschitz with respect to α, and 1-Lipschitz with respect to r, we have

$$\max_{0\le\alpha\le1,0\le r\le R}\left|\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)-\mathbb{E}_{r_t^*}[l_\alpha(r,r_t^*)]\right|\le R\sqrt{\frac{\log(2t/\delta)}{2(t-1)}}+\frac{2R}{\sqrt{t}}.$$

Finally, due to Theorem 1 we have for all α and r,

$$\begin{aligned}\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r_t(\alpha),r_i^*)&\le\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)+\frac{\lambda_t}{1-\lambda_t}\left[\psi(r)-\psi(r_t(\alpha))\right]\\&\le\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)+\frac{1}{\sqrt{t}-1}\max\left\{\psi(0),\psi(R)\right\}\\&\le\frac{1}{t-1}\sum_{i=1}^{t-1} l_\alpha(r,r_i^*)+\frac{R}{\sqrt{t}-1}.\end{aligned}$$

Combining it with the generalization error bound above, with high probability we have for all α and r,

$$\mathbb{E}_{r_t^*}[l_\alpha(r_t(\alpha),r_t^*)]\le\mathbb{E}_{r_t^*}[l_\alpha(r,r_t^*)]+\frac{R}{\sqrt{t}-1}+2\left(R\sqrt{\frac{\log(2t/\delta)}{2(t-1)}}+\frac{2R}{\sqrt{t}}\right).$$

Taking $\min_r$ on the RHS completes the proof. ■

Theorem 5. *With $\lambda_t=1/\sqrt{t}$ and the uniform P_0,* QUANTIZED *achieves* $\mathrm{Regret}_T(\alpha)=O(R\sqrt{T})$.

Proof of Theorem 5: Recall from Section III-C that the quantized true score is denoted by $\tilde{r}_t^*$. From Theorem 2, we have

$$\begin{aligned}\sum_{t=1}^T l_\alpha(r_t(\alpha),\tilde{r}_t^*)-\sum_{t=1}^T l_\alpha(q_\alpha(r_{1:T}^*),\tilde{r}_t^*)&\le\sum_{t=1}^T l_\alpha(r_t(\alpha),\tilde{r}_t^*)-\sum_{t=1}^T l_\alpha(q_\alpha(\tilde{r}_{1:T}^*),\tilde{r}_t^*)\\&=O(R\sqrt{T}).\end{aligned}$$

As $|\tilde{r}_t^*-r_t^*|\le R/\sqrt{T}$ and the quantile loss function $l_\alpha(r,r^*)$ is 1-Lipschitz with respect to r^*, we have

$$\begin{aligned}\left|\sum_{t=1}^T l_\alpha(r_t(\alpha),\tilde{r}_t^*)-\sum_{t=1}^T l_\alpha(r_t(\alpha),r_t^*)\right|&\le\sum_{t=1}^T\left|l_\alpha(r_t(\alpha),\tilde{r}_t^*)-l_\alpha(r_t(\alpha),r_t^*)\right|\\&\le\sum_{t=1}^T|\tilde{r}_t^*-r_t^*|\le R\sqrt{T}.\end{aligned}$$

The comparator term $\sum_{t=1}^T l_\alpha(q_\alpha(r_{1:T}^*),\tilde{r}_t^*)$ can be related similarly to $\sum_{t=1}^T l_\alpha(q_\alpha(r_{1:T}^*),r_t^*)$, and combining the above completes the proof. ■

Theorem 6. *With $\lambda_t=\lambda=\frac{\sqrt{1-\beta}}{\beta+\sqrt{1-\beta}}$ and the uniform P_0, the output $r_t(\alpha)$ of* DISCOUNTED *satisfies*

$$r_t(\alpha)=\arg\min_{r\in\mathbb{R}}\left[(1-\beta)^{-1}\left(\frac{\lambda}{1-\lambda}+\beta^{t-1}\right)\psi(r)+\sum_{i=1}^{t-1}\beta^{t-1-i}l_\alpha(r,r_i^*)\right],$$

for all α and t. In addition, for all $\alpha \in [0,1]$, it guarantees the discounted regret bound

$$\mathrm{Regret}_{T,\beta}(\alpha) := \sum_{t=1}^{T} \beta^{T-t} l_\alpha(r_t(\alpha), r_t^*) - \min_{r\in[0,R]} \sum_{t=1}^{T} \beta^{T-t} l_\alpha(r, r_t^*) \le \frac{R}{\sqrt{1-\beta}} + o(R),$$

where $o(\cdot)$ is with respect to $T \to \infty$.

Proof of Theorem 6: Analogous to Eq.(9), we can write the output of DISCOUNTED as

$$\begin{aligned} r_t(\alpha) &= q_\alpha \left[\lambda P_0 + (1-\lambda)\bar{P}_\beta(r_{1:t-1}^*)\right] \\ &= \min\left\{ r : \frac{r}{R}\left[\lambda + \beta^{t-1}(1-\lambda)\right] + (1-\lambda)(1-\beta)\sum_{i=1}^{t-1} \beta^{t-1-i} \mathbf{1}[r_i^* \le r] \ge \alpha \right\}. \end{aligned}$$

Similar to Eq.(10), this can be verified as a minimizer of the objective

$$H_t(r) := (1-\beta)^{-1}\left(\frac{\lambda\beta^{1-t}}{1-\lambda} + 1\right)\psi(r) + \sum_{i=1}^{t-1} \beta^{-i} l_\alpha(r, r_i^*).$$

For the convenience of notation, we will write the regularizer weight as $z_t := (1-\beta)^{-1}\left(\frac{\lambda\beta^{1-t}}{1-\lambda} + 1\right)$.

Notice that with the uniform P_0, the base regularizer ψ is R^{-1}-strongly-convex due to Theorem 1, therefore we can apply the strong-convexity-based FTRL analysis [22, Corollary 7.9] on the scaled loss functions,

$$h_t(r) := \beta^{-t} l_\alpha(r, r_t^*).$$

This yields

$$\sum_{t=1}^{T} h_t(r_t(\alpha)) - \min_{r\in[0,R]} \sum_{t=1}^{T} h_t(r) \le z_T \left[\max_{r\in[0,R]} \psi(r) - \min_{r\in[0,R]} \psi(r)\right] + \frac{R}{2}\sum_{t=1}^{T} \frac{g_t^2}{z_t^2},$$

where g_t can be any subgradient of $h_t(r)$ at $r = r_t(\alpha)$. Scaling both sides by β^T, we recover the discounted regret definition on the LHS:

$$\mathrm{Regret}_{T,\beta}(\alpha) \le \beta^T z_T \left[\max_{r\in[0,R]} \psi(r) - \min_{r\in[0,R]} \psi(r)\right] + \frac{R\beta^T}{2}\sum_{t=1}^{T} \frac{g_t^2}{z_t}.$$

Next we simplify the obtained expression. The range of ϕ is contained in $[0, R/2]$. In addition, $|g_t| \le \beta^{-t}$ since the quantile loss $l_\alpha(r, r^*)$ is 1-Lipschitz with respect to r. Therefore,

$$\sum_{t=1}^{T} \frac{g_t^2}{z_t} \le \sum_{t=1}^{T} \frac{\beta^{-2t}}{(1-\beta)^{-1}\left(\frac{\lambda\beta^{1-t}}{1-\lambda} + 1\right)} \le \frac{1-\beta}{\beta}\frac{1-\lambda}{\lambda}\sum_{t=1}^{T}\beta^{-t} \le \frac{1-\lambda}{\lambda}\beta^{-T-1},$$

$$\mathrm{Regret}_{T,\beta}(\alpha) \le \frac{R}{2}\left(\frac{\beta^T}{1-\beta} + \frac{\lambda}{1-\lambda}\frac{\beta}{1-\beta} + \frac{1-\lambda}{\lambda}\frac{1}{\beta}\right).$$

Notice that our choice of λ satisfies $\frac{\lambda}{1-\lambda} = \beta^{-1}\sqrt{1-\beta}$, therefore

$$\mathrm{Regret}_{T,\beta}(\alpha) \le \frac{R}{2}\left(\frac{\beta^T}{1-\beta} + \frac{2}{\sqrt{1-\beta}}\right) = \frac{R}{\sqrt{1-\beta}} + o(R),$$

where $o(\cdot)$ is with respect to $T \to \infty$. ■

C. Auxiliary Lemma

Lemma A.1 (Lemma 7.6 of [22]). *Let f be a μ-strongly convex function with respect to a norm $\|\cdot\|$, over a convex set V. For all $x, y \in V$ and subgradients $g \in \partial f(y)$, $g' \in \partial f(x)$, we have*

$$f(x) - f(y) \le \langle g, x-y\rangle + \frac{1}{2\mu}\|g - g'\|_*^2.$$

Here $\langle\cdot,\cdot\rangle$ denotes the inner product, and $\|\cdot\|_*$ denotes the dual norm of $\|\cdot\|$.

The following lemma is a standard tool in ML due to [15].

Lemma A.2 (Hoeffding's inequality)**.** *Let $x_1, \ldots, x_n$ be iid samples of a real-valued random variable on $[a, b]$. Let $\bar{x}$ be the mean of the distribution. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^{n} x_i - \bar{x}\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right).$$

The next lemma is the celebrated Dvoretzky–Kiefer–Wolfowitz inequality, due to [6, 20].

Lemma A.3 (DKW inequality)**.** *Let $x_1, \ldots, x_n$ be iid samples of a real-valued random variable with cumulative distribution function F, and let $\bar{P}(x_{1:n})$ be the empirical distribution of $x_{1:n}$, with cumulative distribution function $\hat{F}_n$. For all $\varepsilon > 0$, we have*

$$\mathbb{P}\left[\sup_{x\in\mathbb{R}}\left|\hat{F}_n(x) - F(x)\right| > \varepsilon\right] \leq 2\exp(-2n\varepsilon^2).$$

D. Experiment

Complementing our theoretical results, we now evaluate the performance of Algorithm 1 using more experiments.

a) Switching Sequence: First, to demonstrate the failure of ERM without the iid assumption, we consider a synthetic $r^*_{1:T}$ sequence which switches in every round between 0 and 1. Similar to Section II, four algorithms are tested: OGD [10], MVP [3],[4] ERM and our Bayesian algorithm QUANTIZED. Figure 3 plots their regret measured by the quantile loss, under two different α values.

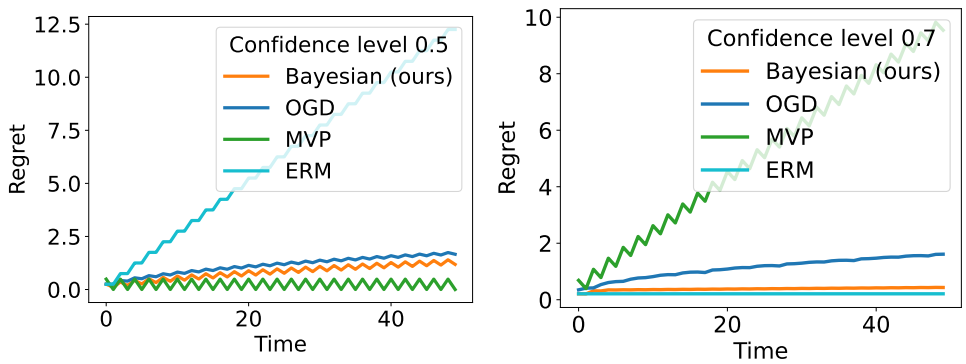

Fig. 3: Regret on switching data.

Consistent with the classical online learning theory, ERM becomes brittle when α matches the long run average of $r^*_{1:T}$ (i.e, 0.5), suffering linear regret with respect to T. In contrast, both OGD (with $\eta_t = t^{-1/2}$; α is known) and our Bayesian algorithm achieve sublinear regret under both α values. Quite different from the conventional online learning framework, MVP is designed to minimize the conditional empirical coverage error, but nonetheless, it achieves low regret when $\alpha = 0.5$. The limitation is that MVP requires a relatively long period to warm up: when $\alpha = 0.7$, the regret of MVP grows linearly at the beginning, before hitting a plateau at $T \approx 800$.

b) Stock Price: Next, we move to an actual CP problem: predicting the time-varying volatility of the stock price, with the base model being a standard time series forecasting method called GARCH [5]. This experiment was designed by [10] and further studied by [3]. See [3, Appendix B.3.1] for the specifics of its context.

Two baselines are considered: a specialization of OGD (ACI) for time series forecasting, and MVP. Besides requiring a fixed learning rate, the former operates on a sliding time window whose length is also a hyperparameter. Similarly, MVP requires picking the size of discretization. For both baselines, we follow the exact implementation from [3], including the hyperparameters.

As for our approach, we adopt the discounted version to handle the continual distribution shift, together with quantization. The discretization grid $\tilde{A}$ has the same size as the MVP baseline, and we pick the discount factor β such that the effective length $(1-\beta^2)^{-1}$ of the discounted time window exactly matches the length of the ACI baseline's sliding window. Given this β, λ_t is selected according to Theorem 6. It means compared to the baselines, our algorithm cannot benefit from any extra hyperparameter tuning.

With $\alpha = 0.9$, Figure 4 plots the $r_{1:T}(\alpha)$ sequence predicted by different algorithms. As a visual sanity check, our algorithm generates a reasonable prediction sequence with slightly less fluctuation than the baselines. To make a more concrete comparison,

[4]Similar to Section II, we use a single random seed for the MVP baseline throughout this section, since we find the results to be generally insensitive to the seed.

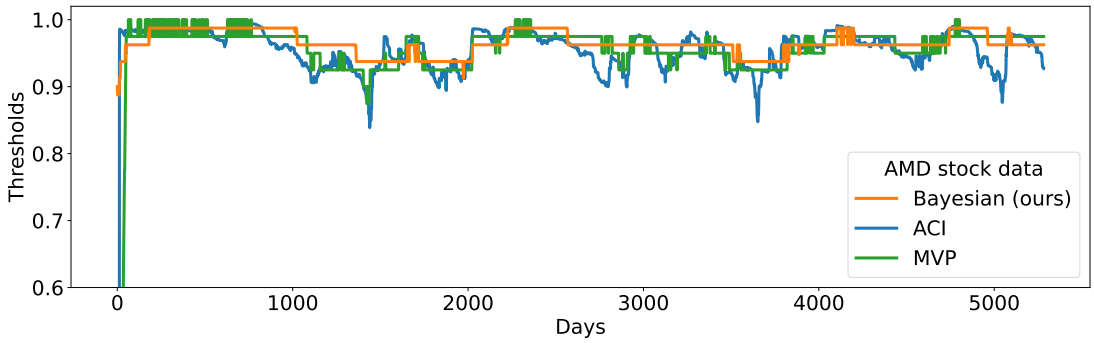

Fig. 4: Predicted score threshold on AMD stock data.

Figure 5 plots the total quantile loss suffered by all three algorithms, as well as the difference compared to ACI. It shows that our algorithm achieves almost the same total loss as ACI, and it is faster to warm up than MVP.

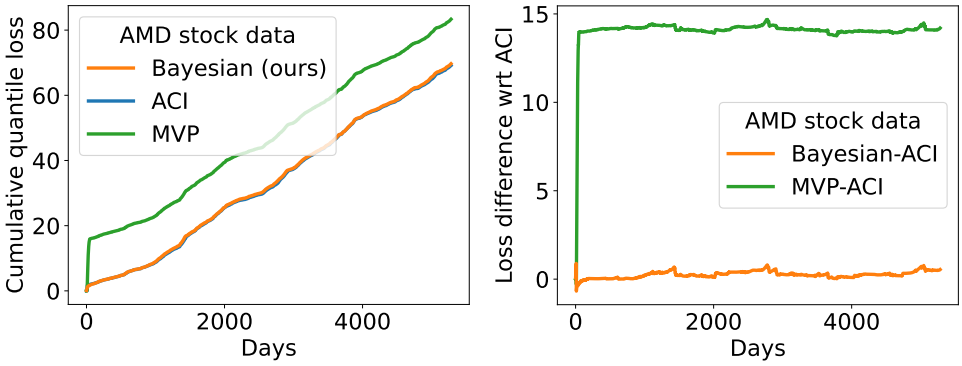

Fig. 5: Quantile loss on AMD stock data.

We also evaluate the empirical coverage rate of the tested algorithms. Although our algorithm is not designed for this metric, it performs competitively compared to the baselines. The target is $1 - \alpha = 0.9$, and closer to this target is better. ACI achieves 0.901, MVP achieves 0.893, and our Bayesian algorithm achieves 0.899.

Below are the results using a different dataset (NVDA instead of AMD). Figure 6 plots the predicted thresholds, and Figure 7 plots the total quantile loss.

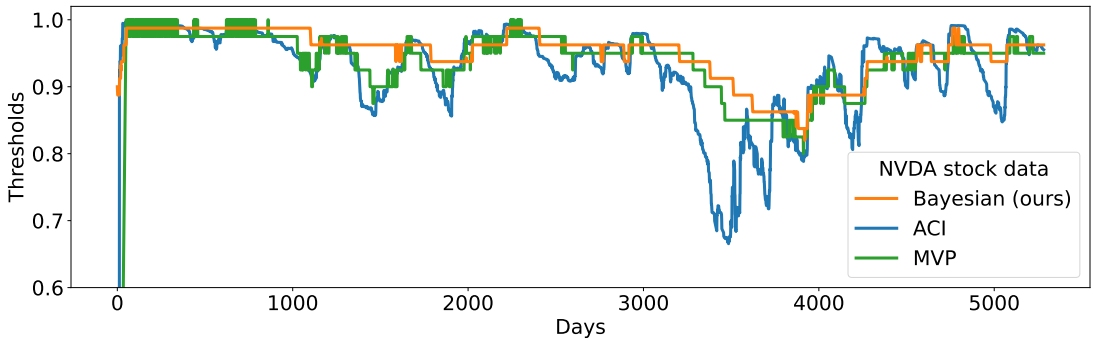

Fig. 6: Predicted score threshold on NVDA stock data.

As for the coverage frequency, ACI achieves 0.899, MVP achieves 0.891, and our Bayesian algorithm achieves 0.897. Again, closer to the target 0.9 is better. The conclusion is that in the fixed-α setting our algorithm performs competitively compared to the baselines, while in the multi-α setting it demonstrates the advantage from Section II, i.e., not suffering from the monotonicity issue.

c) Better Prior: Finally, we demonstrate that better empirical performance can be obtained by using a better prior.

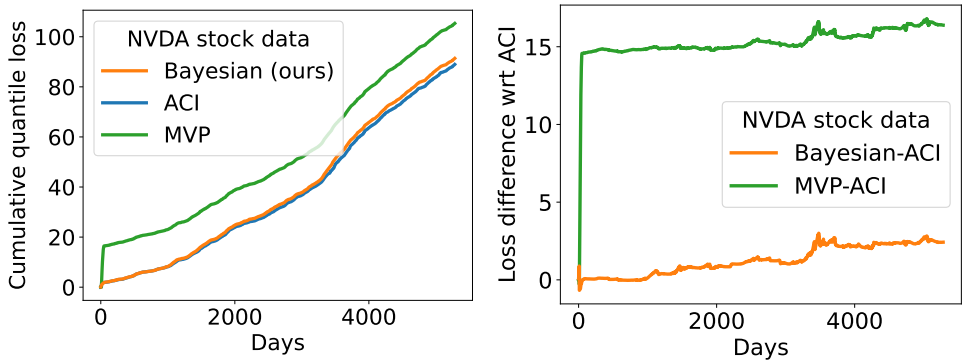

Fig. 7: Quantile loss on NVDA stock data.

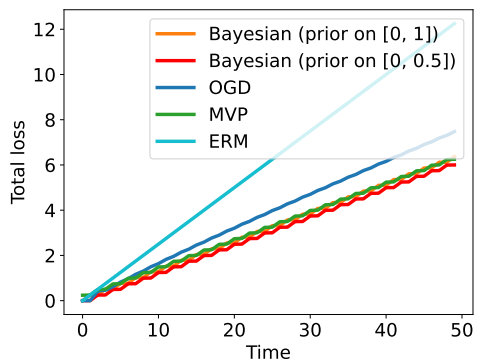

Fig. 8: Demonstrating the role of prior.

To this end, we consider a synthetic, switching data experiment similar to Figure 3. Here the true scores sequence is set as the switching sequence 0.5, 0, 0.5, 0, ..., against which ERM is known to fail when predicting the median (i.e., has large regret). We test two versions of our algorithm; the first version selects the prior as Uniform([0,1]), while the second version selects the prior as Uniform([0, 0.5]). Slightly different from our Figure 3, we evaluate the total quantile loss as the performance metric (lower is better), while the rest of the experimental setup (including the hyperparameters of the baselines) remains the same.

Results are shown in Figure 8. The observation is that with the default Uniform([0,1]) prior, our algorithm (orange) performs similarly as the MVP baseline while beating both the OGD and the ERM baseline. With the Uniform([0, 0.5]) prior which is better tailored to the range of the data, our algorithm (red) obtains additional improvement in its performance.