Trading-Off Payments and Accuracy in Online Classification with Paid Stochastic Experts

Dirk van der Hoeven [*1] **Ciara Pike-Burke** [*2] **Hao Qiu** [3] **Nicolò Cesa-Bianchi** [3 4]

Abstract

We investigate online classification with paid stochastic experts. Here, before making their prediction, each expert must be paid. The amount that we pay each expert directly influences the accuracy of their prediction through some unknown Lipschitz "productivity" function. In each round, the learner must decide how much to pay each expert and then make a prediction. They incur a cost equal to a weighted sum of the prediction error and upfront payments for all experts. We introduce an online learning algorithm whose total cost after T rounds exceeds that of a predictor which knows the productivity of all experts in advance by at most $\mathcal{O}\big(K^2(\ln T)\sqrt{T}\big)$ where K is the number of experts. In order to achieve this result, we combine Lipschitz bandits and online classification with surrogate losses. These tools allow us to improve upon the bound of order $T^{2/3}$ one would obtain in the standard Lipschitz bandit setting. Our algorithm is empirically evaluated on synthetic data.

1. Introduction

We investigate online classification in the framework of prediction with expert advice where, in each round, the learning agent predicts an unknown binary label by aggregating the stochastic predictions of a number of experts. At the end of each round, the learner observes the true label and updates the function used to aggregate experts. In the variant considered in this work, we assume that at the beginning of a round the learner allocates a payment to each expert which affects the expert's performance in that round. This payment model of expert advice is realistic in many scenarios since human annotators will often only give useful advice if they are adequately compensated, and machine annotators may require more computation to return accurate predictions. Moreover, monetary incentives have been studied in crowdsourcing (Ho et al., 2015; 2016). Although this is a different setting to that considered here, it is natural to study the effect of these payments in online binary classification with stochastic expert advice.

Motivated by results in crowdsourcing—e.g., (Ho et al., 2016)—we assume that each expert has a *productivity function* which determines the probability that they predict the label correctly given the payment they received. The productivity function can be different for each expert and is initially unknown to the learner. In each round, the learner pays each expert $j = 1, \dots, K$ some amount $c_j \in [0, 1]$ before observing their advice. The accuracy of the advice that expert j returns depends on the amount they are paid through the unknown productivity function, $p_j : [0, 1] \to [0, 1]$, where $p_j(c)$ is the probability that expert j is correct when the payment is c. The learner can use the expert advice to improve their prediction, but they also want to minimize the payments to the experts. Therefore, they must trade-off between the price of the expert advice, and any improvements to prediction accuracy it may bring.

We define the learner's cost over a sequence of T rounds as the sum of classification mistakes and payments to experts. If the probabilities $p_j(c_j)$ are known for some $c_1, \dots, c_K$ and for all experts $j = 1, \dots, K$, then we can write down the Bayes-optimal cost. In particular, if the events that each expert makes a mistake are independent, then the error probability of the Bayes-optimal aggregated prediction is known to decrease exponentially fast with exponent $-2\big(\gamma_1^2(c_1) + \cdots + \gamma_K^2(c_K)\big)$, where $\gamma_j(c_j) = \big(p_j(c_j) - \frac{1}{2}\big)$ for $j = 1, \dots, K$ (Ho et al., 2013). Given productivity functions $\boldsymbol{p} = \big(p_1, \dots, p_K\big)$, we then define the optimal cost over T rounds by $T\,\mathrm{OPT}(\boldsymbol{p})$, where

$$\mathrm{OPT}(\boldsymbol{p}) = \min_{c_1,\dots,c_K \in [0,1]} \left\{ e^{-2\sum_{j=1}^K \gamma_j(c_j)^2} + \lambda \sum_{j=1}^K c_j \right\}$$

and $\lambda > 0$ is a given parameter introduced to balance the

[*]Equal contribution [1]Korteweg-de Vries Institute for Mathematics University of Amsterdam, Amsterdam, The Netherlands [2]Department of Mathematics, Imperial College London, London, UK [3]Università degli Studi di Milano, Milan, Italy [4]Politecnico di Milano, Milan, Italy. Correspondence to: Dirk van der Hoeven <dirk@dirkvanderhoeven.com>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

trade-off between accuracy and payments.

In this paper, we consider the case when $\boldsymbol{p}$ is unknown and the learner's goal is to minimize the regret,

$$R_T = \sum_{t=1}^{T} \left(\mathbb{P}_Z\big(\widehat{y}_t \neq y_t\big) + \lambda \sum_{j=1}^{K} c_{t,j} \right) - T\,\textsc{Opt}(\boldsymbol{p}) \quad (1)$$

where, in each round t, $\widehat{y}_t \in \{-1, 1\}$ is the learner's prediction (which depends on the stochastic expert advice Z), $y_t \in \{-1, 1\}$ is the adversarially chosen true label, and $c_{t,j}$ is the payment for expert j. Following the standard online learning protocol, we assume the true label y_t is revealed at the end of each round. The learner can then use y_t and the expert advice to learn the productivity functions.

Regret minimization in this problem presents several challenges. The first is the need to trade-off cost and accuracy while simultaneously learning how the payments affect the accuracy. Indeed, we only observe the predictions of the experts with the payments that we pay them. This introduces an additional exploration-vs-exploitation trade-off as is typically seen in bandit problems. However, as we discuss in Section 2, this is more challenging than in standard bandit problems since the relationship between the payments that we choose and the regret is more complex and possibly non-smooth.

A further significant challenge is to combine the predictions from all experts when we only have estimates of their accuracy. In particular, if we have estimated an experts productivity function $\widehat{p}_{t,j}(c_j)$, as being close to 0 or 1 for a specific c_j in round t, directly using a majority or weighted majority aggregation approach could lead to undesirable scaling of the regret with $1/\min\{\widehat{p}_{t,j}(c_j), 1 - \widehat{p}_{t,j}(c_j)\}$ which can be arbitrarily large.

Our approach to overcome these issues is based on a combination of several ideas. First, we discretize the payment interval and for each payment and expert combination we estimate the probability of a correct classification. To deal with the exploration vs exploitation challenge we rely on the optimism in the face of uncertainty principle—see, for example, (Lattimore & Szepesvári, 2020, Chapter 7). While this is a standard approach for stochastic losses, if it were to be directly applied to the discretized payments here, it would lead to an undesirable regret bound of $\mathcal{O}(T^{2/3})$.[1] Instead, we combine the optimistic principle with tools from online classification with surrogate losses to obtain a $\mathcal{O}(\sqrt{T})$ regret bound. Specifically, we use the randomized predictions of Van der Hoeven (2020); Van der Hoeven et al. (2021) which gains us considerable leeway in the analysis, see Section 4 for the details.

To avoid regret that scales with $1/\min\{p_j(c), 1 - p_j(c)\}$, we propose a modified aggregation approach. This approach simply follows the advice of one expert if we believe they are very likely to be correct (or wrong, in which case we use the opposite of their prediction). This aggregation approach allows us to have *multiplicative* control over estimation errors rather than additive control. Combined with the randomized prediction technique, multiplicative control over the estimation error is a crucial element in our analysis, and one of the major differences compared to standard analysis of aggregated classifiers.

Combining these ideas with tight error bounds on the productivity function allows us to obtain the following result (implied by Theorem 4.7).

Theorem 1.1. *The regret of LCB-GAPTRON (Algorithm 1) satisfies $R_T = \mathcal{O}\big(K^2(\ln T)\sqrt{T}\big)$ where K is the number of experts and T is the number of rounds.*

This result represents an improvement on the $T^{2/3}$ regret bound that would be achievable if we were to simply use an optimistic algorithm with discretized costs. We also demonstrate that our algorithm significantly outperforms this naive algorithm in several simulated environments. Our experiments also indicate that the most computationally demanding step of our algorithm can be replaced by a cheaper approximation with little impact on the regret.

1.1. Related Work

Online aggregation of experts is also studied in online boosting (Chen et al., 2012; Beygelzimer et al., 2015), a setting where there are no payments and the predictions of experts are adversarial. When the average accuracy of the experts is unknown, Beygelzimer et al. (2015) prove a mistake bound of $\frac{8}{\gamma^2 K}T + \tilde{\mathcal{O}}\big(\frac{K}{\gamma^2}\big)$, where $\frac{1}{2} - \gamma$ is an upper bound on the fraction of mistakes made by any expert. Note that, due to the adversarial assumption on the experts, the leading term in this bound vanishes at rate K^{-1}, as opposed to the exponential rate e^{-K} achievable in our stochastic setting.

Our setting is also related to the framework of online prediction with limited expert advice (Seldin et al., 2014; Kale, 2014), where the predictions of experts and the payments are both chosen by an adversary. In this model, the learner can buy advice from any subset of experts at the price posted by the adversary. As the payments are not chosen by the learner, the trade-off between payments and accuracy is different from the one studied in this work.

Although our setting is online learning, solving classification tasks by aggregating the predictions of stochastic

[1] To see why standard methods give a $T^{2/3}$ rate consider a simplified setting with one expert. Then, the mistake probability is equivalent to the productivity function of that expert, denoted by p. Since we assume that p is Lipschitz, the problem is now reduced to Lipschitz bandits, for which a $T^{2/3}$ bound is known to be unavoidable (Slivkins et al., 2019).

experts naturally arises also in the context of crowdsourced labeling (Zhang et al., 2016; Vaughan, 2017). Karger et al. (2014) study a setting where a requester has a set of n homogeneous binary labeling tasks to be assigned to m workers arriving in random order. Each worker j is characterized by an unknown probability p_j of labeling correctly any task and by a capacity T_j. At each round, the requester observes the capacity of the current worker and selects a subset of tasks respecting the worker's capacity. After the m workers have been all assigned tasks, the requester aggregates their predictions to infer a single prediction for each task.

This model has been extended to consider heterogeneous tasks (so that each worker j has a different accuracy $p_{i,j}$ for each task i) and costly access to ground truth labels for each task type (Ho et al., 2013). Tran-Thanh et al. (2013) also extend the crowdsourcing model to a setting where workers have fixed and known costs and the requester must allocate tasks while respecting a budget constraint.

From an online learning viewpoint, these crowdsourcing papers consider a dual *pool-based* model, where a set of unlabeled points is preliminary given and experts arrive online. In contrast, in our problem the set of experts is fixed and new instances are considered in each round, thus the problem studied here is quite distinct from crowdsourcing. In our work, we also consider payments that influence the workers' accuracy, which is not included in the classical crowdsourcing model.

Monetary incentives in crowdsourcing have been considered by (Ho et al., 2015; 2016; Shah & Zhou, 2016), and although the crowdsourcing setting is distinct from ours, these works help motivate our setting of paid stochastic experts. Ho et al. (2015) empirically showed the effect of monetary incentives on the quality of the predictions in crowdsourcing. Ho et al. (2016) introduce an online stochastic model where workers, who act strategically, are drawn i.i.d. from a fixed and unknown distribution of worker types. Each type determines the workers' productivity function and the workers' effort function, where the latter controls their strategic behavior. Because of strategic behaviors, their payment scheme is more complex than ours. On the other hand, in their model the requester's utility cannot be increased by aggregating workers on the same task.

Another example of strategic workers is investigated by Shah & Zhou (2016), where they compute minimal payments sufficient to incentivize workers to predict labels that they are sure of, and abstain on the others.

2. Preliminaries

In each round $t \in [T]$ of online classification with paid experts, the learner chooses a payment $c_{t,j} \in [0,1]$ for each expert $j \in [K]$. After receiving the payments, the experts reveal their predictions $Z_{t,1}, \dots, Z_{t,K} \in \{-1,1\}^K$ for the true label $y_t \in \{-1,1\}$. For each $j \in [K]$, the prediction $Z_{t,j}$ is stochastic and satisfies

$$\mathbb{P}(Z_{t,j} \neq y_t) = p_j(c_{t,j})$$

where p_j is the productivity function of expert j. Based on the expert advice, the learner then predicts $\widehat{y}_t \in \{-1,1\}$, observes y_t, and suffers the zero-one loss $\mathbb{1}[\widehat{y}_t \neq y_t]$. The sequence $y_1, y_2, \dots$ of true labels is arbitrary and deterministic[2], and the events $\{Z_{t,j} \neq y_t\}$ for $t \in [T]$ and $j \in [K]$ are stochastic and assumed to be independent. We let $Z_t = (Z_{t,1}, \dots, Z_{t,K})$ be the vector of all experts' predictions in round t.

We assume the experts' productivity functions $p_1, \dots, p_K : [0,1] \to [0,1]$ are L-Lipschitz,

$$\left|p_j(c) - p_j(c')\right| \leq L|c - c'| \qquad c, c' \in [0,1]\,. \tag{2}$$

This class of productivity functions is broad enough to capture most realistic settings, including the special cases where the productivity funtion is monotonic, logistic, or where the experts are restricted to predict the correct label with probability greater than 0.5. These productivity functions are initially unknown to the learner. For any round $t = 1, \dots, T$, we define the filtration

$$\mathcal{F}_t = \sigma(Z_1, B_1, y_1, \dots, Z_{t-1}, B_{t-1}, y_{t-1})$$

where B_s represents any internal randomization used by the learner in round $s \leq t$. The learner can use any information in $\mathcal{F}_t$ to estimate the productivity functions, decide on the payments and aggregate expert predictions in round t. The learners objective is to select payments and aggregate expert advice to minimize the cumulative regret defined in (1).

To understand the ideas behind the algorithm and the involved challenges, we first consider a simplified setting where payments do not affect the prediction accuracy. In other words, $p_j(c) = p'_j$ for all $c \in [0,1]$, for all $j \in [K]$, and for some $p'_1, \dots, p'_K \in [0,1]^K$. As argued in the introduction, if $p'_1, \dots, p'_K$ are known, then the learner's expected number of mistakes is exponentially decreasing in K. To see this, let $w(p') = \frac{1}{2} \ln \frac{p'}{1-p'}$ and assume the learner's prediction is

$$\widehat{y}_t = \operatorname{sign}\left(\sum_{j=1}^{K} w(p'_j) Z_{t,j}\right)\,. \tag{3}$$

Similarly to the analysis of AdaBoost (Schapire & Freund, 2013), we can upper bound the zero-one loss with the expo-

[2] Our results continue to hold even when the labels are stochastic, provided that the events $\{Z_{t,j} \neq y_t\}$ remain independent.

nential loss and write

$$\mathbb{P}_{Z_t}(\widehat{y}_t \neq y_t) \leq \mathbb{E}_{Z_t}\left[\exp\left(-y_t \sum_{j=1}^{K} w(p'_j) Z_{t,j}\right)\right]$$
$$= \prod_{j=1}^{K}\left(p'_j\sqrt{\frac{1-p'_j}{p'_j}} + (1-p'_j)\sqrt{\frac{p'_j}{1-p'_j}}\right) \qquad (4)$$
$$= \prod_{j=1}^{K}\sqrt{4\big(1-(\tfrac{1}{2}-p'_j)^2\big)} \leq \exp\left(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p'_j)^2\right).$$

The first equality holds because of our choice of $w(p')$ and the assumption that the predictions of experts are independent. The last inequality uses $1 + x \leq e^x$.

The tightness of the bound (4) is easily verified in the special case $(\frac{1}{2} - p'_j)^2 = \gamma^2$ for all $j \in [K]$. Then $\widehat{y}_t$ is the majority vote, which is clearly Bayes optimal and, assuming K is odd,

$$\mathbb{P}_Z(\widehat{y}_t \neq y_t) = \mathrm{Binom}\left(K, \left\lfloor \frac{K}{2} \right\rfloor, \frac{1}{2} + \gamma\right)$$
$$= \Omega\left(\sqrt{\frac{1}{K}}\, e^{\frac{K}{2}\ln(1-4\gamma^2)}\right)$$

where $\mathrm{Binom}(n, m, p)$ is the probability of at most m heads in n flips of a coin with bias p, see Ferrante (2021, inequality (15)). This implies that, in the worst case, we are suboptimal by a factor $K^{-1/2}$ as soon as we apply the first inequality in (4).

We now explain our approach for learning the unknown probabilities p'_j. Let $\widehat{p}_{t,j}$ be the estimate of p'_j in round t. Following the derivation of (4) and using prediction (3) with estimates $\widehat{p}_j$ in lieu of the true probabilities p'_j, we see that

$$\mathbb{P}_{Z_t}(\widehat{y}_t \neq y_t)$$
$$\leq \prod_{j=1}^{K}\left(p'_j\sqrt{\frac{1-\widehat{p}_j}{\widehat{p}_j}} + (1-p'_j)\sqrt{\frac{\widehat{p}_j}{1-\widehat{p}_j}}\right) \qquad (5)$$

A first challenge is to control the difference between (4) and (5). This involves controlling terms of order

$$\sqrt{(1-\widehat{p}_j)/\widehat{p}_j} - \sqrt{(1-p'_j)/p'_j}\,.$$

Via standard online learning analysis, we would obtain a regret bound scaling linearly with the Lipschitz constant of the function $\sqrt{(1-p')/p'}$, which is of order $1/p'$. But this would require enforcing that $\min_j p'_j$ be bounded away from 0 (and from 1 for the symmetric function $\sqrt{p'/(1-p')}$).

A second challenge is learning the optimal cost for each expert. This is a bandit problem over continuously many actions, because choosing a payment $c \in [0, 1]$ for some expert does not provide any information about payments $c' \neq c$. Using Lipschitzness of the productivity functions, we can discretize the payments. However, as we argued above, the key function $\sqrt{(1-p')/p'}$, which controls the error estimating the probability of mistake, is not Lipschitz in $[0, 1]$. This necessitates further algorithmic developments and novel analyses. In the following section we introduce our algorithm and explain how we overcome the aforementioned challenges.

3. Algorithm

Our algorithm LCB-GAPTRON for online classification with paid experts is presented in Algorithm 1. At a high level, our algorithm selects payments using the pessimistic principle (since we receive bandit feedback for the payments), and then uses a randomized weighted aggregation procedure to predict a label based on the expert advice. However, as indicated above, several adjustments need to be made to account for the intricacies of the problem and ensure that we can obtain a $\mathcal{O}(\sqrt{T})$ regret bound. We detail these below.

LCB-GAPTRON requires as input a discrete set $\mathcal{C}$ of payments. To learn the optimal payment in $\mathcal{C}$ for each expert, it is helpful to maintain an empirical estimate

$$\widehat{p}_{t+1,j}(c) = \sum_{s=1}^{t} \mathbb{1}[c_{s,j} = c]\frac{1 + Z_{s,j}y_t}{2n_{t+1,j}(c)}$$

of the probability of success for each $c \in \mathcal{C}$, where $n_{t+1,j}(c) = \sum_{s=1}^{t} \mathbb{1}[c_{s,j} = c]$ is the number of times we have paid expert j c up to the end of round t. We then construct optimistic estimates $\mathcal{P}_t(c)$ in line 1 based on the empirical Bernstein bound (Audibert et al., 2007)—see Lemma A.1 in the Appendix—which are used when computing the payments to the experts in (6).

The algorithm also requires a parameter $\beta \geq 0$, whose role is to control the cutoff value $\alpha_{t,j}(c)$ of the estimated probabilities for each $c \in \mathcal{C}$. As discussed around (5), a key challenge in this problem is that the function that we optimize is subject to large changes when the estimated probabilities are close to 0 or 1. To solve this problem, when $p_{t,j}(c)$ is very large or very small, we simply follow experts j that we estimate are either very good or very bad (in the latter case, the weight $w_{t,j}$ is negative). However, this does not resolve our troubles completely. Indeed, standard online methods would still suffer regret inversely proportional to the cutoff value because they need to control the difference between the estimated probabilities and the true probabilities. To overcome this issue, we show that we can estimate the true probabilities up to a multiplicative factor of $\frac{3}{2}$. This is also a result of using the empirical Bernstein bound to construct the confidence intervals. Indeed, the empirical Bernstein

Algorithm 1 LCB-GAPTRON

Require: Set $\mathcal{C}$ of N costs, parameters $\beta \geq 0$ and $\delta > 0$

Initialize: For $j = 1, \ldots, K$ and $c \in \mathcal{C}$, set $\widehat{p}_{1,j}(c) = 1$, $\mathcal{P}_{1,j}(c) = 1$, $\alpha_{1,j}(c) = \frac{1}{2}$, and $n_{1,j}(c) = 0$
for $t = 1 \ldots T$ **do**
  if $t \leq N$ **then**
    Find $c_{t,1}, \ldots, c_{t,K}$ such that $c_{t,k} = 0$ for all $k \in [K]$.
  else
    Compute $c_{t,1}, \ldots, c_{t,K}$ solution of

$$\operatorname*{arg\,min}_{c_1,\ldots,c_K \in \mathcal{C}^K} e^{-2\sum_{j=1}^K \left(\frac{1}{2} - \mathcal{P}_{t,j}(c_j)\right)^2} + \lambda \sum_{j=1}^K c_j \tag{6}$$

  end if
  Pay $c_{t,1}, \ldots, c_{t,K}$ and receive advice $Z_{t,1}, \ldots Z_{t,K}$
  if some $\widehat{p}_{t,j}(c_{t,j}) \notin \left[\alpha_{t,j}(c_{t,j}), 1 - \alpha_{t,j}(c_{t,j})\right]$ **then**
    Pick any j such that

$$\widehat{p}_{t,j}(c_{t,j}) \notin \left[\alpha_{t,j}(c_{t,j}), 1 - \alpha_{t,j}(c_{t,j})\right]$$

    Predict $\widehat{y}_t = \operatorname{sign}(\widehat{p}_{t,j}(c_{t,j}) - \frac{1}{2}) Z_{t,j}$
  else
    for $j = 1, \ldots, K$ **do**
      $w_{t,j}(\widehat{p}_{t,j}(c_{t,j})) \leftarrow \frac{1}{2} \ln \frac{\widehat{p}_{t,j}(c_{t,j})}{1 - \widehat{p}_{t,j}(c_{t,j})}$
    end for
    Set $x_t = \sum_{j=1}^K w_{t,j}(\widehat{p}_{t,j}(c_{t,j})) Z_{t,j}$
    Predict

$$\widehat{y}_t = \begin{cases} \operatorname{sign}(x_t) & \text{w.p. } 1 - \frac{1}{2} e^{-x_t \operatorname{sign}(x_t)} \\ -\operatorname{sign}(x_t) & \text{w.p. } \frac{1}{2} e^{-x_t \operatorname{sign}(x_t)} \end{cases} \tag{7}$$

  end if
  Obtain true label y_t
  for $j = 1, \ldots, K$ **do**
    $n_{t+1,j}(c_{t,j}) \leftarrow n_{t,j}(c_{t,j}) + 1$
    $\alpha_{t+1,j}(c_{t,j}) \leftarrow \min\left\{\frac{\beta}{n_{t+1,j}(c_{t,j})}, \frac{1}{2}\right\}$
    $\widehat{p}_{t+1,j}(c_{t,j}) \leftarrow \sum_{s=1}^t \mathbb{1}[c_{s,j} = c_{t,j}] \frac{1 + Z_{s,j} y_t}{2 n_{t+1,j}(c_{t,j})}$
    $s_{t+1}(c_{t,j}) \leftarrow \operatorname{sign}\left(\frac{1}{2} - \widehat{p}_{t+1,j}(c_{t,j})\right)$
    $Q_{t+1}(c_{t,j}) \leftarrow \min\left\{1 - \widehat{p}_{t+1,j}(c_{t,j}), \widehat{p}_{t+1,j}(c_{t,j}), \frac{3\ln(3/\delta)}{n_{t+1,j}(c_{t,j})} + \sqrt{\frac{\widehat{p}_{t+1,j}(c_{t,j})(1 - \widehat{p}_{t+1,j}(c_{t,j}))}{n_{t+1,j}(c_{t,j})} 2\ln(\frac{3}{\delta})}\right\}$
    $\mathcal{P}_{t+1,j}(c_{t,j}) \leftarrow \widehat{p}_{t+1,j} - s_{t+1}(c_{t,j}) Q_{t+1}(c_{t,j})$
  end for
end for

bound allows us to have both additive and multiplicative control of the estimated probabilities which is essential for our analysis. See Section 4 for further details.

To avoid suffering additional regret for only being able to estimate the probabilities up to a multiplicative factor, we use the ideas of Van der Hoeven (2020); Van der Hoeven et al. (2021). In particular, whenever all the estimated probabilities are bounded away from 0 and 1, we use randomized predictions. These randomized predictions $\widehat{y}_t = \widehat{y}_t(B_t, Z_t)$ defined in (7)—where B_t is the internal randomization used by Algorithm 1 at step t—satisfy (see Lemma 4.2)

$$\mathbb{P}_{B_t}\left(\widehat{y}_t \neq y_t \mid Z_t, \mathcal{F}_t\right) \leq \frac{1}{2} e^{-y_t \sum_{j=1}^K w_{t,j}(\widehat{p}_{t,j}(c_{t,j})) Z_{t,j}}$$

where $c_{t,j} \in \mathcal{C}$ is the payment to expert j in round t and $\widehat{p}_{t,j}(c_{t,j})$ is the estimate of the accuracy of expert j when the payment is $c_{t,j}$. This gains us a factor $\frac{1}{2}$ compared to the bound we used in Section 2 (see equation (4)) and allows us to compensate for the multiplicative factor introduced by estimating the probabilities.

4. Analysis

In this section we prove the regret bounds in the introduction. To condense notation slightly, let

$$\Phi\big(\boldsymbol{p}(\boldsymbol{c})\big) = e^{-2\sum_{j=1}^K \left(\frac{1}{2} - p_j(c_j)\right)^2}$$

where $\boldsymbol{p}(\boldsymbol{c})$ is the vector with elements $p_j(c_j)$. Then, $\textsc{opt}(\boldsymbol{p}) = \min_{\boldsymbol{c} \in [0,1]^K} \Phi\big(\boldsymbol{p}(\boldsymbol{c})\big) + \lambda \sum_{j=1}^K c_j$. Our proof of the regret bound follows from the below steps.

Step 1: Bound the error in estimated probabilities We start by showing that we can control the difference between true probabilities $p_j(c)$ and estimated probabilities $\widehat{p}_{t,j}(c)$. For any $\delta \in (0,1)$ and $t \geq 1$, let $\Lambda_{\delta,t}$ be the event that

$$\left|\widehat{p}_{t,j}(c) - p_j(c)\right| \leq \frac{3}{n_{t,j}(c)} \ln \frac{3}{\delta} + \sqrt{\frac{\widehat{p}_{t,j}(c)(1 - \widehat{p}_{t,j}(c))}{n_{t,j}(c)} 2 \ln \frac{3}{\delta}} \tag{8}$$

simultaneously for all $j \in [K]$ and $c \in \mathcal{C}$.

Lemma 4.1. *Fix $\delta \in (0,1)$. Suppose that $t > N$. Then $\Lambda_{\delta,t}$ holds with probability at least $1 - \delta T N K$.*

The proof of Lemma 4.1 follows from an application of the union bound and the empirical Bernstein bound (Audibert et al., 2007) and is provided in the Appendix.

Step 2: Bound the prediction error We now turn our attention to controlling the number of mistakes we make. In the following, we bound the probability that we make a mistake in any given round t. We split the analysis into two

cases: either all $\widehat{p}_{t,j}(c_{t,j}) \in \big[\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})\big]$, or at least one $\widehat{p}_{t,j}(c_{t,j})$ is outside of the specified range (meaning that some expert is certain so we can predict just using their advice). We denote by $\mathcal{E}_t$ the event that in round t we are in the first case, that is that $\widehat{p}_{t,j}(c_{t,j}) \in \big[\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})\big]$ for all j.

As a first step, observe that when $\mathcal{E}_t$ holds we issue a randomized prediction, see equation (7) in Algorithm 1. This prediction has the following crucial property.

Lemma 4.2. *For any round $t > N$, assuming $\mathcal{E}_t$ holds, the prediction in equation (7) satisfies*

$$P_{B_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t, Z_t) \leq \tfrac{1}{2} e^{-y_t \sum_{j=1}^K w_{t,j}(\widehat{p}_{t,j}(c_{t,j}))Z_{t,j}}.$$

The proof of Lemma 4.2 can be found in the appendix and essentially follows from Van der Hoeven et al. (2021, Lemma 1) although our proof is simpler.

As a second step, let us integrate out the randomness in Z_t in the bound in Lemma 4.2. Using the definition of $w_{t,j}$ we obtain:

$$P_{B_t,Z_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t) \leq \frac{1}{2}\prod_{j=1}^K \left(p_j(c_{t,j})\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + (1-p_j(c_{t,j}))\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}}\right). \tag{9}$$

Now, under the assumptions that all $\widehat{p}_{t,j}(c_{t,j}) \in \big[\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})\big]$ and that $\Lambda_{\delta,t}$ holds, for all $\beta > 0$

$$\big|\widehat{p}_{t,j}(c_{t,j}) - p_j(c_{t,j})\big| = \widetilde{\mathcal{O}}\left(\frac{1}{\beta}\widehat{p}_{t,j}(c_{t,j})\right).$$

To see why this holds, observe that $\widehat{p}_{t,j}(c_{t,j}) \geq \alpha_{t,j}(c_{t,j}) = \mathcal{O}\big(\frac{\beta}{n_{t,j}(c_{t,j})}\big)$, which we can use to bound the right-hand side of the equation in the definition of $\Lambda_{\delta,t}$ in equation (8). We can then substitute this into equation (9) and, after some manipulation together with a careful choice of β, show that when $\mathcal{E}_t$ and $\Lambda_{\delta,t}$ both hold,

$$\mathbb{P}_{B_t,Z_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t) \leq \frac{3}{4}\prod_{j=1}^K \sqrt{1-4\big(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j})\big)^2}$$

after which, using $1+x \leq \exp(x)$, we arrive at the statement of Lemma 4.3.

Lemma 4.3. *Let $\beta = 18\ln(3/\delta)K^2$. Then for any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ and $\mathcal{E}_t$ both hold,*

$$\mathbb{P}_{B_t,Z_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t) \leq \frac{3}{4}\Phi\big(\widehat{\boldsymbol{p}}_t(\boldsymbol{c}_t)\big)$$

Step 3: Relate $\Phi\big(\widehat{\boldsymbol{p}}(\boldsymbol{c})\big)$ to $\Phi\big(\boldsymbol{p}(\boldsymbol{c})\big)$ In the next lemma we show how to control the difference between the right-hand side of Lemma 4.3 and the same equation, but with $p_j(c_{t,j})$ instead of $\widehat{p}_{t,j}(c_{t,j})$.

Lemma 4.4. *For any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ holds,*

$$\frac{3}{4}\Phi\big(\widehat{\boldsymbol{p}}_t(\boldsymbol{c}_t)\big) \leq \frac{7}{8}\Phi\big(\boldsymbol{p}(\boldsymbol{c}_t)\big) + 96K\sum_{j=1}^K\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \left(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\right)^2\right).$$

To prove Lemma 4.4 (see the Appendix for a detailed proof) we show that for some constant $b > 0$ and for all $\boldsymbol{q}, \boldsymbol{q}' \in [0,1]^K$,

$$\Phi(\boldsymbol{q}) - \Phi(\boldsymbol{q}') \leq \tfrac{1}{8}\Phi(\boldsymbol{q}) + bK\sum_{j=1}^K |q_j - q'_j|^2.$$

This means that

$$\tfrac{3}{4}\Phi(\boldsymbol{q}) = \tfrac{7}{8}\Phi(\boldsymbol{q}) - \tfrac{1}{8}\Phi(\boldsymbol{q}) \leq \frac{7}{8}\left(\Phi(\boldsymbol{q}') + bK\sum_{j=1}^K |q_j - q'_j|^2\right).$$

Since we assumed that the event $\Lambda_{\delta,t}$ holds, we can replace q_j and q'_j by $\widehat{p}_{t,j}(c_{t,j})$ and $p_j(c_{t,j})$ to prove Lemma 4.4.

Step 4: Bound the loss from pessimistic choice of payments Next, we need to control the difference between paying the experts our chosen costs versus the optimal costs. Using the optimism in the face of uncertainty principle, the same ideas we used to prove Lemma 4.4, and the Lipschitz assumption in equation (2) we arrive at the following lemma, whose proof can be found in the Appendix.

Lemma 4.5. *Let $\mathcal{C}$ be such that for any $c^\star \in [0,1]$ there is a $\widetilde{c} \in \mathcal{C}$ that satisfies $|\widetilde{c} - c^\star| \leq \varepsilon$. Then for any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ holds,*

$$\frac{7}{8}\Phi\big(\boldsymbol{p}(\boldsymbol{c}_t)\big) + \lambda\sum_{j=1}^K c_{t,j} \leq \min_{\boldsymbol{c}\in[0,1]^K}\left\{\Phi\big(\boldsymbol{p}(\boldsymbol{c})\big) + \lambda\sum_{j=1}^K c_j\right\} + (4L+\lambda)K\varepsilon + 96K\sum_{j=1}^K\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \left(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\right)^2\right).$$

Step 5: Control the regret in when estimated probabilities are too large or small Combined, Lemmas 4.1, 4.3, and 4.4 give us control over the regret in rounds where all $\widehat{p}_{t,j}(c_{t,j}) \in \big[\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})\big]$. In the case where there is at least one $\widehat{p}_{t,j}(c_{t,j})$ which is not in the above range, then we can control the regret by using the following observation. In rounds where $\mathcal{E}_t$ does not hold we follow the (flipped) prediction of an expert j that satisfies

$\widehat{p}_{t,j}(c_{t,j}) \notin \big[\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})\big]$. For simplicity suppose that we just follow the actual prediction of that expert, which implies that $\widehat{p}_{t,j}(c_{t,j}) \leq \alpha_{t,j}(c_{t,j})$. In this case $\widehat{y}_t$ does not depend on B_t and we have that, assuming $\Lambda_{\delta,t}$ and $\mathcal{E}_t^c$ both hold,

$$\begin{aligned}\mathbb{P}_{Z_t}(\widehat{y}_t \neq y_t \mid \mathcal{F}_t) &= p_j(c_{t,j}) \\ &\leq \widehat{p}_{t,j}(c_{t,j}) + |p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j})| \\ &= \widetilde{\mathcal{O}}(\alpha_{t,j}(c_{t,j})) = \widetilde{\mathcal{O}}\Big(\frac{\beta}{n_{t,j}(c_{t,j})}\Big)\,,\end{aligned}$$

where the third equality follows from the assumption that events $\Lambda_{\delta,t}$ and $\mathcal{E}_t^c$ hold. To see why, observe that since $\widehat{p}_{t,j}(c_{t,j}) \leq \alpha_{t,j}(c_{t,j})$ equation (8) is of order $\alpha_{t,j}(c_{t,j})$. The formal result can be found in Lemma 4.6 below, the full proof of which can be found in the Appendix.

Lemma 4.6. *For any round $t > N$ and $\delta \in (0,1)$, assuming $\mathcal{E}_t^c$ and $\Lambda_{\delta,t}$ both hold,*

$$\begin{aligned}&\mathbb{P}_{Z_t}\big(\widehat{y}_t \neq y_t \mid \mathcal{F}_t\big) + \lambda \sum_{j=1}^{K} c_{t,j} \\ &\leq \textsc{opt}(\boldsymbol{p}) + \frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + (4L+\lambda)K\varepsilon \\ &\quad + 96K \sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right).\end{aligned}$$

Step 6: Sum up the per-round regret We now have a per-round control of the cost and probability of making a mistake. Thus, by combining Lemmas 4.1, 4.3, 4.4, 4.5, and 4.6 and summing over rounds we arrive at the main result of the paper in Theorem 4.7 below, whose result is implied by Theorem A.2 in the Appendix. As stated below, the choice of $\mathcal{C}$ in Algorithm 1 that delivers the bounds of Theorem 4.7 is the uniform ε-grid on $[0,1]$.

Theorem 4.7. *Let $\mathcal{C}$ be such that for any $c^\star \in [0,1]$ there is a $\widetilde{c} \in \mathcal{C}$ that satisfies $|\widetilde{c} - c^\star| \leq \varepsilon$, let $\beta = 18\ln(3/\delta)K^2$, and let $\delta = \frac{1}{(1+\lambda K)T^2K}$. Then the regret of Algorithm 1 satisfies*

$$R_T = \mathcal{O}\left(\varepsilon KT(4L+\lambda K) + N\Big(\lambda + K^3\ln(\lambda KT)^2\Big)\right)\,.$$

Theorem 4.7 implies that if $N = \mathcal{O}(\varepsilon^{-1})$ for ε of order $\sqrt{\big(\lambda + K^3\ln(\lambda KT)^2\big)/\big(KT(4L+\lambda K)\big)}$, then our final bound is

$$R_T = \mathcal{O}\left(K^2\sqrt{T(4L+\lambda)\Big(\lambda + \ln(1\lambda KT)^2\Big)}\right). \quad (10)$$

A slight modification of the proof shows that if we compete with the best costs in $\mathcal{C}$, rather than the best costs in $[0,1]^K$, then the corresponding regret,

$$\begin{aligned}R_T(\mathcal{C}) = \mathbb{E}\left[\sum_{t=1}^{T}\Big(\mathbb{1}[\widehat{y}_t \neq y_t] + \lambda\sum_{j=1}^{K} c_{t,j}\Big)\right] \\ - T\min_{\boldsymbol{c}\in\mathcal{C}}\left\{\Phi\big(\boldsymbol{p}(\boldsymbol{c})\big) + \lambda\sum_{j=1}^{K} c_j\right\}\end{aligned}$$

can be bounded by $\mathcal{O}\Big(N\big(\lambda + K^3\ln(\lambda KT)^2\big)\Big)$. Note that in this case, we remove the first term of the regret in Theorem 4.7 as there is no discretization error. Moreover, if we have a discrete set of costs $\mathcal{C}$, then we do not require the costs to be Lipschitz.

5. Alternative Approaches

In this section we formulate alternative notions of regret that make the problem of online classification with paid experts amenable to solution via standard bandit algorithms that predict using the advice of a single expert. We start by considering a finite set $\mathcal{C}$ of costs and assume that in each round the learner must pick a single expert and a cost $c \in \mathcal{C}$ to pay them. Assuming $p_j(c) \geq \frac{1}{2}$ for all experts $j \in [K]$ and costs $c \in \mathcal{C}$, we may define the regret

$$\begin{aligned}R_T^{\text{band}} = \mathbb{E}\left[\sum_{t=1}^{T}\Big(\mathbb{1}[\widehat{y}_t \neq y_t] + \lambda c_t\Big)\right] \\ - T\min_{j\in[K], c\in\mathcal{C}}\big(p_j(c) + \lambda c\big)\,.\end{aligned}$$

R_T^{band} presents a different trade-off between costs and mistakes. In particular, while the term accounting for the costs is considerably smaller (as only one expert is paid in each round), the term accounting for the expected number of mistakes is exponentially larger (in the number of experts) because there is no expert aggregation. Treating each expert and cost pair as an arm, we can use LCB (the variant of the UCB algorithm using lower confidence bounds to minimize losses instead of maximizing rewards) and immediately obtain the bound $R_T^{\text{band}} = \widetilde{\mathcal{O}}\big(\sqrt{KNT}\big)$.

When the productivity functions are L-Lipschitz, see (2), we can define a harder notion of regret R_T^{cont} in which the comparator is defined with respect to the best cost $c \in [0,1]$, as opposed to the best $c \in \mathcal{C}$ which we used in the definition of R_T^{band}. Now, if we discretize the interval $[0,1]$ using the grid $\mathcal{C}$ of Theorem 4.7 with $|\mathcal{C}| = \mathcal{O}\big(\varepsilon^{-1}\big)$, then we pay an approximation error of $\varepsilon T(L+\lambda)$. By running LCB on $\mathcal{C}\times K$ arms, and adding the approximation error, we get, after tuning ε, $R_T^{\text{cont}} = \mathcal{O}\big(T^{2/3}(KL)^{1/3}\big)$.

Although these bounds on R_T^{band} and R_T^{cont} are not directly comparable to the bounds on our different notion of regret (1), in the next section we perform an empirical com-

parison between variants of Algorithm 1 and the instance of LCB run on $K \times N$ arms which we used to bound R_T^{band}.

6. Experiments

Our experiments use two sets of productivity functions defined on a uniform (random) grid $\mathcal{C}$ of N payments on $\left[\frac{1}{2}, 1\right]$. The first productivity function is linear with the same slope for all experts: We sampled N numbers $c_1, \dots, c_N$ uniformly at random from $\left[\frac{1}{2}, 1\right]$ and defined $p_j(c_i) = c_i$ for all $i \in [N]$ and all $j \in [K]$. The second productivity function is sigmoidal, with a different slope for each expert: We sampled N numbers $c_1, \dots, c_N$ uniformly at random from $[0, 1]$ and K integers $\theta_1, \dots, \theta_K$ uniformly at random from $[1, 10]$. For each $j \in [K]$ we then set

$$p_j(c_i) = \frac{\exp(\theta_j c_i)}{1 + \exp(\theta_j c_i)} \,. \tag{11}$$

A fresh sample of the productivity functions for each expert is drawn in each repetition of our experiments.

Consistent with our definition of regret, we measure the performance of the algorithms in terms of their cost; i.e., the total number of mistakes plus the total payments to the experts.

The running time for finding the optimal costs in T rounds using our optimistic estimates (6) in LCB-GAPTRON (Algorithm 1) is $\mathcal{O}(TN^K)$, which prevents running experiments for moderately large values of the parameters. Therefore, we designed two simple and efficient approximations of the optimal payments defined in (6). The first one, SELFISH, optimizes the cost for each expert independently of the others. The second one, LOCAL, computes the optimal cost of each expert iteratively, in a round-robin fashion, while keeping the cost of the other experts fixed. We call BRUTE the inefficient implementation of LCB-GAPTRON that directly optimizes (6) using brute force search. Finally, we call LCB the instance of LCB run over $K \times N$ actions which we defined in Section 5.

In our first experiment we pick sufficiently small values for the parameters so that we can run LCB-GAPTRON-BRUTE. Figure 1 shows that for both choices of the productivity function, all instances of LCB-GAPTRON have nearly indistinguishable performance and outperform LCB. Thus, we can safely drop LCB-GAPTRON-BRUTE and run a second set of experiments using a larger time horizon $T = 10^5$.

In our second experiment, we test LCB-GAPTRON-SELFISH and LCB-GAPTRON-LOCAL against LCB using the second productivity function (plots using the first productivity function are similar, see Appendix B). The results in Figure 2 show that for small value of the scaling parameters ($K = 10$ and $N = 10$) and large cost units ($\lambda = 10^{-2}$), LCB is eventually on par with LCB-GAPTRON. Recall that

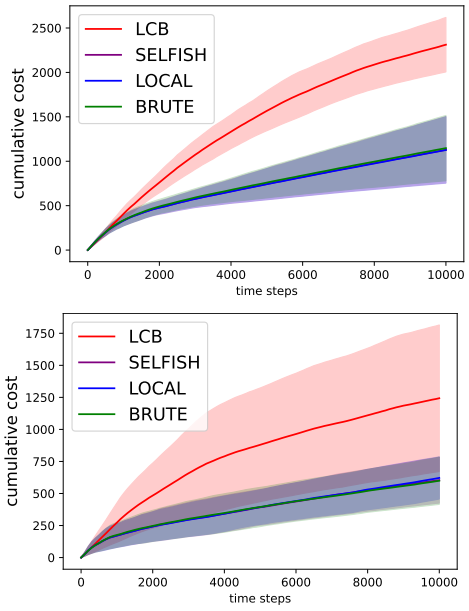

Figure 1. Cumulative cost over time for the choice of parameters $K = 5$, $N = 5$, $T = 10^4$, $\lambda = 10^{-2}$. The algorithms are run using the first productivity function in the top plot and the second productivity function in the bottom plot. The error bars show the standard deviation of the cost averaged over 20 repetitions.

LCB pays a single expert at each round, which is the reason that LCB catches up with LCB-GAPTRON. However, for larger values of the scaling parameters ($K = 20$ and $N = 50$) or for smaller values of the cost units ($\lambda = 10^{-3}$), both variants of LCB-GAPTRON dominate again.

In Appendix B, we report the performance of LCB, LCB-GAPTRON-SELFISH, and LCB-GAPTRON-LOCAL on both productivity functions for $T = 10^5$, $K \in \{10, 20\}$, $N \in \{10, 50, 100\}$, and $\lambda \in \left\{10^{-2}, 10^{-3}\right\}$. These plots essentially confirm the observations made in the discussion of Figure 2. In Appendix B we also provide the figures for the experiments with LCB-GAPTRON-BRUTE with $\lambda \in \left\{10^{-2}, 10^{-3}\right\}$, $N = 5$, $K = 5$, $T = 10^5$, and both productivity functions, which also confirm the observations made in the discussion of Figure 1.

7. Future Work

In this paper we have studied online classification with paid stochastic experts, and presented an algorithm which has sub-linear regret in terms of the number of rounds T. We have also demonstrated that this algorithm performs well in several experimental settings. Although the algorithm is computationally expensive, we have shown empirically that using approximations does not lead to much deterioration of performance. Several questions remain open. For example, whether a computationally efficient algorithm with similar

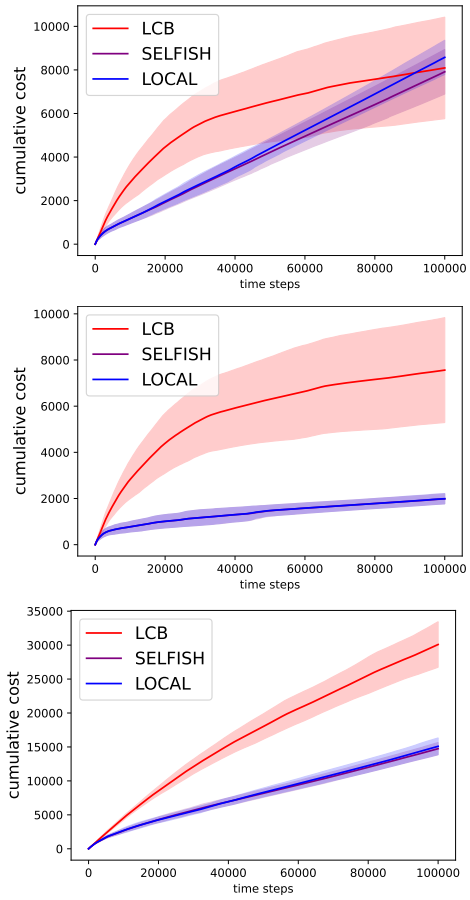

Figure 2. Total cost for the following choices of parameters: $K = 10, N = 10, T = 10^5, \lambda = 10^{-2}$ (top plot), $K = 10, N = 10, T = 10^5, \lambda = 10^{-3}$ (central plot), $K = 20, N = 50, T = 10^5, \lambda = 10^{-2}$ (bottom plot). The algorithms are run using the second productivity function. The error bars show the standard deviation of the cost averaged over 20 repetitions.

regret guarantees can be developed, and whether the K^2 factor in our regret bound $K^2(\ln T)\sqrt{T}$ can be improved on. On the other hand, we conjecture the $\sqrt{T}$ rate is optimal, because of the need of estimating the discretized productivity functions from bandit feedback.

There is scope to extend our model in several directions by considering, for example, strategic experts (as in Roughgarden & Schrijvers (2017); Freeman et al. (2020)), or making the experts' performance to depend also on contextual information. It is also an open question whether faster rates can be achieved with stronger parametric assumptions on the productivity function. For example, if the productivity function is a sigmoid, $p_j(c) = \exp(a + bc)/(1 + \exp(a + bc))$ with unknown constants $a, b, \in \mathbb{R}$, can the regret be significantly improved?

Acknowledgements

This work was mostly done while DvdH was at the University of Milan partially supported by the MIUR PRIN grant Algorithms, Games, and Digital Markets (ALGADIMAR) and partially supported by Netherlands Organization for Scientific Research (NWO), grant number VI.Vidi.192.095. HQ and NCB are partially supported by the MIUR PRIN grant Algorithms, Games, and Digital Markets (ALGADIMAR), by the EU Horizon 2020 ICT-48 research and innovation action under grant agreement 951847, project ELISE (European Learning and Intelligent Systems Excellence), and by the FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, investment 1.3, line on Artificial Intelligence). CPB gratefully acknowledges the support of the Imperial College European Partners Fund.

References

Audibert, J.-Y., Munos, R., and Szepesvári, C. Tuning bandit algorithms in stochastic environments. In *International conference on algorithmic learning theory*, pp. 150–165. Springer, 2007.

Beygelzimer, A., Kale, S., and Luo, H. Optimal and adaptive algorithms for online boosting. In *International Conference on Machine Learning*, pp. 2323–2331. PMLR, 2015.

Chen, S.-T., Lin, H.-T., and Lu, C.-J. An online boosting algorithm with theoretical justifications. *arXiv preprint arXiv:1206.6422*, 2012.

Ferrante, G. C. Bounds on binomial tails with applications. *IEEE Transactions on Information Theory*, 67(12):8273–8279, 2021.

Freeman, R., Pennock, D. M., Podimata, C., and Vaughan, J. W. No-regret and incentive-compatible prediction with expert advice. *arXiv preprint arXiv:2002.08837*, 2020.

Ho, C.-J., Jabbari, S., and Vaughan, J. W. Adaptive task assignment for crowdsourced classification. In *International Conference on Machine Learning*, pp. 534–542. PMLR, 2013.

Ho, C.-J., Slivkins, A., Suri, S., and Vaughan, J. W. Incentivizing high quality crowdwork. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 419–429, 2015.

Ho, C.-J., Slivkins, A., and Vaughan, J. W. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. *Journal of Artificial Intelligence Research*, 55:317–359, 2016.

Van der Hoeven, D. Exploiting the surrogate gap in online multiclass classification. *Advances in Neural Information Processing Systems*, 2020.

Van der Hoeven, D., Fusco, F., and Cesa-Bianchi, N. Beyond bandit feedback in online multiclass classification. In *Advances in neural information processing systems*, 2021.

Kale, S. Multiarmed bandits with limited expert advice. In *Conference on Learning Theory*, pp. 107–122. PMLR, 2014.

Karger, D. R., Oh, S., and Shah, D. Budget-optimal task allocation for reliable crowdsourcing systems. *Operations Research*, 62(1):1–24, 2014.

Lattimore, T. and Szepesvári, C. *Bandit algorithms*. Cambridge University Press, 2020.

Roughgarden, T. and Schrijvers, O. Online prediction with selfish experts. *Advances in Neural Information Processing Systems*, 30, 2017.

Schapire, R. E. and Freund, Y. Boosting: Foundations and algorithms. *Kybernetes*, 2013.

Seldin, Y., Bartlett, P., Crammer, K., and Abbasi-Yadkori, Y. Prediction with limited advice and multiarmed bandits with paid observations. In *International Conference on Machine Learning*, pp. 280–287. PMLR, 2014.

Shah, N. B. and Zhou, D. Double or nothing: Multiplicative incentive mechanisms for crowdsourcing. *Journal of Machine Learning Research*, 17:1–52, 2016.

Slivkins, A. et al. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

Tran-Thanh, L., Venanzi, M., Rogers, A., and Jennings, N. R. Efficient budget allocation with accuracy guarantees for crowdsourcing classification tasks. In *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, pp. 901–908, 2013.

Vaughan, J. W. Making better use of the crowd: How crowdsourcing can advance machine learning research. *J. Mach. Learn. Res.*, 18(1):7026–7071, 2017.

Zhang, J., Wu, X., and Sheng, V. S. Learning from crowdsourced labeled data: a survey. *Artificial Intelligence Review*, 46(4):543–576, 2016.

A. PROOF DETAILS

Lemma A.1 (Audibert et al. (2007)). *Let $X_1, \ldots, X_t$ be i.i.d. random variables with $\mathbb{E}[X] = \mu$ taking their values in $[0,1]$. Denote by $\bar{X}_t = \frac{1}{t}\sum_{s=1}^{t} X_s$ and by $V_t = \frac{1}{t}\sum_{s=1}^{t}(X_s - \bar{X}_t)^2$. Then with probability at least $1-\delta$*

$$|\bar{X}_t - \mu| \leq \sqrt{\frac{V_t}{t} 2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{t}.$$

Lemma 4.2. *For any round $t > N$, assuming $\mathcal{E}_t$ holds, the prediction in equation (7) satisfies*

$$P_{B_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t, Z_t) \leq \tfrac{1}{2} e^{-y_t \sum_{j=1}^{K} w_{t,j}(\widehat{p}_{t,j}(c_{t,j}))Z_{t,j}}.$$

Proof. We denote by $x_t = \sum_{j=1}^{K} w_{t,j}(\widehat{p}_{t,j}(c_{t,j}))Z_{t,j}$ The proof also follows from Lemma 1 by Van der Hoeven et al. (2021). Suppose that $\text{sign}(x_t) = \text{sign}(y_t)$. Then $\mathbb{E}[\mathbb{1}[\widehat{y}_t \neq y_t]|\mathcal{F}_t, Z_t] = \frac{1}{2}\exp(-y_t x_t)$. Now suppose that $\text{sign}(x_t) \neq \text{sign}(y_t)$. Then

$$\mathbb{E}_{\widehat{y}_t}[\mathbb{1}[\widehat{y}_t \neq y_t]|\mathcal{F}_t, Z_t] = 1 - \tfrac{1}{2}\exp(y_t x_t) = 1 + \tfrac{1}{2}\exp(-y_t x_t) - \tfrac{1}{2}\exp(y_t x_t) - \tfrac{1}{2}\exp(-y_t x_t) \leq \tfrac{1}{2}\exp(-y_t x_t),$$

where the last inequality is due to $\frac{1}{2}\exp(y_t x_t) + \frac{1}{2}\exp(-y_t x_t) \geq 1$, which holds by Jensen's inequality. □

Lemma 4.1. *Fix $\delta \in (0,1)$. Suppose that $t > N$. Then $\Lambda_{\delta,t}$ holds with probability at least $1 - \delta TNK$.*

Proof. First, note that since $t > N$ $n_t(c) \geq 1$ for all $c \in \mathcal{C}$. Denote by $V_{t,j}(c) = \frac{1}{|\mathcal{N}_t|}\sum_{i\in\mathcal{N}_t}\big(\sum_{i'\in\mathcal{N}_t}\frac{1+Z_{i',j}y_t}{2|\mathcal{N}_t|} - \frac{1+Z_{i,j}y_t}{2}\big)^2$, where $\mathcal{N}_t = \{i < t : \mathbb{1}[c_{i,j} = c]\}$ and denote by

$$\begin{aligned}\widehat{V}_{t,j}(c) &= \frac{1}{n_{t,j}(c)}\sum_{s=1}^{t-1}\mathbb{1}[c_{s,j}=c]\left(\sum_{s'=1}^{t-1}\mathbb{1}[c_{s',j}=c]\frac{1+Z_{s',j}y_t}{n_{t,j}(c)} - \frac{1+Z_{s,j}(c_{s,j})y_t}{2}\right)^2 \\ &= \widehat{p}_{t,j}(c)(1-\widehat{p}_{t,j}(c))\end{aligned}$$

Now, by reindexing the sum we can see that

$$\begin{aligned}&\frac{1}{n_{t,j}(c)}\sum_{s=1}^{t-1}\mathbb{1}[c_{s,j}=c]\left(\sum_{s'=1}^{t-1}\mathbb{1}[c_{s',j}=c]\frac{1+Z_{s',j}y_t}{n_{t,j}(c)} - \frac{1+Z_{s,j}(c_{s,j})y_t}{2}\right)^2 \\ &= \frac{1}{|\mathcal{N}_t|}\sum_{i\in\mathcal{N}_t}\big(\sum_{i'\in\mathcal{N}_t}\frac{1+Z_{i',j}y_t}{2|\mathcal{N}_t|} - \frac{1+Z_{i,j}y_t}{2}\big)^2.\end{aligned}$$

Therefore, we can see that

$$\begin{aligned}&P\left(|\widehat{p}_{t,j}(c) - p_j(c)| \geq \sqrt{\frac{\widehat{V}_{t,j}(c)}{n_{t,j}(c)}2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{n_{t,j}(c)}\right) \\ &\leq P\left(\exists\,\mathcal{N}_h \in \{\mathcal{N}_{N+1}, \ldots, \mathcal{N}_t\} \quad s.t. \quad |\sum_{i\in\mathcal{N}_h}\frac{1+Z_{i,j}(c)y_t}{2|\mathcal{N}_h|} - p_j(c)| \geq \sqrt{\frac{V_{h,j}(c)}{|\mathcal{N}_h|}2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{|\mathcal{N}_h|}\right) \\ &\leq \sum_{h=N+1}^{t} P\left(|\sum_{i\in\mathcal{N}_h}\frac{1+Z_{i,j}(c)y_t}{2|\mathcal{N}_h|} - p_j(c)| \geq \sqrt{\frac{V_{n,j}(c)}{|\mathcal{N}_h|}2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{|\mathcal{N}_h|}\right) \leq T\delta,\end{aligned}$$

where the last inequality is due to Lemma A.1. Taking the union bound we can see that the above holds for all j and $c \in \mathcal{C}$ with probability at most $TNK\delta$. □

Lemma 4.3. *Let $\beta = 18\ln(3/\delta)K^2$. Then for any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ and $\mathcal{E}_t$ both hold,*

$$\mathbb{P}_{B_t, Z_t}(\widehat{y}_t \neq y_t | \mathcal{F}_t) \leq \frac{3}{4}\Phi\big(\widehat{\boldsymbol{p}}_t(\boldsymbol{c}_t)\big)$$

Proof. Since both $\Lambda_{\delta,t}$ and $\mathcal{E}_t$ hold we may use Lemma 4.2 to obtain

$$\begin{aligned}
&\mathbb{E}_Z[\mathbb{E}_{\widehat{y}_t}[\mathbb{1}[\widehat{y}_t \neq y_t]]] \leq \tfrac{1}{2}\mathbb{E}_Z[\exp(-y_t x_t)] \\
&= \tfrac{1}{2}\prod_{j=1}^{K}\left(P(Z_{t,j}(c_{t,j})y_t = 1)\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + P(Z_{t,j}(c_{t,j})y_t = -1)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}}\right) \\
&= \tfrac{1}{2}\prod_{j=1}^{K}\left(p_j(c_{t,j})\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + (1-p_j(c_{t,j}))\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}}\right) \\
&= \tfrac{1}{2}\prod_{j=1}^{K}\Bigg(2\sqrt{\widehat{p}_{t,j}(c_{t,j})(1-\widehat{p}_{t,j}(c_{t,j}))} \\
&\quad + \big(p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j})\big)\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + \big(\widehat{p}_{t,j}(c_{t,j}) - p_j(c_{t,j})\big)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}}\Bigg).
\end{aligned}$$

Recall that that $\alpha_{t,j}(c_{t,j}) = \min\left\{\frac{\beta}{n_{t,j}(c_{t,j})}, \frac{1}{2}\right\}$. This implies that if $\alpha_{t,j}(c_{t,j}) = \frac{1}{2}$ event $\mathcal{E}_t$ can only hold if $\widehat{p}_{t,j}(c_{t,j}) = \frac{1}{2}$, in which case we trivially have

$$\big(p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j})\big)\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + \big(\widehat{p}_{t,j}(c_{t,j}) - p_j(c_{t,j})\big)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}} = 0$$

We do the remainder of the analysis under the assumptions that $\alpha_{t,j}(c_{t,j}) \neq \frac{1}{2}$ and $p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j}) < 0$. The case where $p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j}) \geq 0$ follows from symmetrical arguments. We have that

$$\begin{aligned}
&\big(p_j(c_{t,j}) - \widehat{p}_{t,j}(c_{t,j})\big)\sqrt{\frac{1-\widehat{p}_{t,j}(c_{t,j})}{\widehat{p}_{t,j}(c_{t,j})}} + \big(\widehat{p}_{t,j}(c_{t,j}) - p_j(c_{t,j})\big)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}} \\
&\leq \big(\widehat{p}_{t,j}(c_{t,j}) - p_j(c_{t,j})\big)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}} \\
&\leq \left(\sqrt{\frac{\widehat{V}_{t,j}(c_{t,j})}{n_{t,j}(c_{t,j})}2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\right)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}},
\end{aligned}$$

where the last inequality holds because of the assumption that $\Lambda_{\delta,t}$ holds. Since $\widehat{p}_{t,j}(c_{t,j}) \in [\alpha_{t,j}(c_{t,j}), 1-\alpha_{t,j}(c_{t,j})]$ we have that

$$\begin{aligned}
&\left(\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})(1-\widehat{p}_{t,j}(c_{t,j}))}{n_{t,j}(c_{t,j})}2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\right)\sqrt{\frac{\widehat{p}_{t,j}(c_{t,j})}{1-\widehat{p}_{t,j}(c_{t,j})}} \\
&\leq \big(\sqrt{\frac{2}{\beta}\ln(3/\delta)} + \frac{3}{\beta}\ln(3/\delta)\big)\sqrt{\widehat{p}_{t,j}(c_{t,j})(1-\widehat{p}_{t,j}(c_{t,j}))}
\end{aligned}$$

where we used that $\alpha_{t,j}(c_{t,j}) \leq \frac{\beta}{n_{t,j}(c_{t,j})} \leq 1 - \widehat{p}$. The above implies that

$$\mathbb{E}_Z[\mathbb{E}_{\widehat{y}_t}[\mathbb{1}[\widehat{y}_t \neq y_t]]] \leq \tfrac{1}{2}\Big(1 + \tfrac{1}{2}\big(\sqrt{\frac{2}{\beta}\ln(3/\delta)} + \frac{3}{\beta}\ln(3/\delta)\big)\Big)^K \prod_{j=1}^{K} 2\sqrt{\widehat{p}_{t,j}(c_{t,j})(1-\widehat{p}_{t,j}(c_{t,j}))}$$

Since $\beta = 18\ln(3/\delta)K^2$ we have that

$$\Big(1 + \tfrac{1}{2}\big(\sqrt{\frac{2}{\beta}\ln(3/\delta)} + \frac{3}{\beta}\ln(3/\delta)\big)\Big)^K \leq (1 + \frac{1}{3K})^K \leq 1/(1-1/3) = \frac{3}{2}$$

and thus

$$\mathbb{E}_Z[\mathbb{E}_{\widehat{y}_t}[\mathbb{1}[\widehat{y}_t \neq y_t]]] \leq \frac{3}{4}\prod_{j=1}^{K} 2\sqrt{\widehat{p}_{t,j}(c_{t,j})(1-\widehat{p}_{t,j}(c_{t,j}))}$$
$$= \frac{3}{4}\prod_{j=1}^{K}(1-4(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2) \leq \frac{3}{4}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2\Big)$$

where the last inequality is due to $1+x \leq \exp(x)$. □

Lemma 4.4. *For any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ holds,*

$$\frac{3}{4}\Phi\big(\widehat{\boldsymbol{p}}_t(\boldsymbol{c}_t)\big) \leq \frac{7}{8}\Phi\big(\boldsymbol{p}(\boldsymbol{c}_t)\big) + 96K\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \left(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\right)^2\right).$$

Proof. Denote by $g_t = \exp\Big(-2\sum_{j=1}^{K}(\frac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2\Big)$. Using the fact that $\exp(x)$ and $(\frac{1}{2}-x)^2$ are convex in x we can see that

$$\begin{aligned}
&\frac{3}{4}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2\Big)\\
&= \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2\Big) - \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) - \frac{1}{8}g_t\\
&\leq \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + \frac{28}{8}g_t\sum_{j=1}^{K}|p_j(c_{t,j})-\widehat{p}_{t,j}(c_{t,j})| - \frac{1}{8}g_t\\
&\leq \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + g_t^2/12 + 3\Big(\frac{28}{8}\Big)^2K\sum_{j=1}^{K}|p_j(c_{t,j})-\widehat{p}_{t,j}(c_{t,j})|^2 - \frac{1}{8}g_t\\
&\leq \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + 3\Big(\frac{28}{8}\Big)^2K\sum_{j=1}^{K}|p_j(c_{t,j})-\widehat{p}_{t,j}(c_{t,j})|^2,
\end{aligned}$$

where the third inequality is because $ab \leq \frac{a^2}{2\eta} + \frac{\eta}{2}b^2$ for $\eta > 0$ and the last inequality is due to the fact that $g_t \geq g_t^2$. Therefore, we have that

$$\begin{aligned}
&\frac{3}{4}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-\widehat{p}_{t,j}(c_{t,j}))^2\Big)\\
&\leq \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + 3\Big(\frac{28}{8}\Big)^2K\sum_{j=1}^{K}|p_j(c_{t,j})-\widehat{p}_{t,j}(c_{t,j})|^2\\
&\leq \frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2}-p_j(c_{t,j}))^2\Big) + 6\Big(\frac{28}{8}\Big)^2K\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right),
\end{aligned}$$

where the last inequality is due to the assumption that $\Lambda_{\delta,t}$ holds.

□

Lemma 4.5. *Let $\mathcal{C}$ be such that for any $c^\star \in [0,1]$ there is a $\widetilde{c} \in \mathcal{C}$ that satisfies $|\widetilde{c} - c^\star| \le \varepsilon$. Then for any round $t > N$ and $\delta \in (0,1)$, assuming $\Lambda_{\delta,t}$ holds,*

$$\begin{aligned}
&\frac{7}{8}\Phi\big(\boldsymbol{p}(\boldsymbol{c}_t)\big) + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\le \min_{\boldsymbol{c}\in[0,1]^K} \left\{ \Phi\big(\boldsymbol{p}(\boldsymbol{c})\big) + \lambda \sum_{j=1}^{K} c_j \right\} + (4L+\lambda)K\varepsilon \\
&\quad + 96K \sum_{j=1}^{K} \left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \left(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})} \right)^2 \right) .
\end{aligned}$$

Proof. Since we assume that $\Lambda_{\delta,t}$ holds we have that

$$\min_{\boldsymbol{c}\in\mathcal{C}} \left\{ \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_j))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \right\} \ge \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - \mathcal{P}_{t,j}(c_{t,j}))^2 \Big) + \lambda \sum_{j=1}^{K} c_{t,j}.$$

Denote by $h_t = \exp\Big(-2\sum_{j=1}^{K} (\frac{1}{2} - p_j(c_{t,j}))^2 \Big)$ By the above inequality and the fact that $\exp(x)$ and $(\frac{1}{2} - x)^2$ are convex, we have that

$$\begin{aligned}
&\exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_{t,j}))^2 \Big) + \lambda \sum_{j=1}^{K} c_j - \min_{\boldsymbol{c}\in\mathcal{C}} \left\{ \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_j))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \right\} \\
&\le \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_{t,j}))^2 \Big) - \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - \mathcal{P}_{t,j}(c_{t,j}))^2 \Big) \\
&\le 4h_t \sum_{j=1}^{K} |\mathcal{P}_{t,j}(c_{t,j}) - p_j(c_{t,j})| \qquad (12) \\
&\le \frac{1}{8}h_t^2 + 48K \sum_{j=1}^{K} |\mathcal{P}_{t,j}(c_{t,j}) - p_j(c_{t,j})|^2 \\
&\le \frac{1}{8}h_t^2 + 96K \sum_{j=1}^{K} \left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})} \Big)^2 \right)
\end{aligned}$$

where the third inequality is because $ab \le \frac{a^2}{2\eta} + \frac{\eta}{2}b^2$ for $\eta > 0$ and the last inequality is due to the assumption that $\Lambda_{\delta,t}$ holds. Thus, we have that

$$\begin{aligned}
&\frac{7}{8}\exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_{t,j}))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \\
&= \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_{t,j}))^2 \Big) + \lambda \sum_{j=1}^{K} c_j - h_t \\
&\le \min_{\boldsymbol{c}\in\mathcal{C}} \left\{ \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_j))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \right\} + 96K \sum_{j=1}^{K} \left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})} \Big)^2 \right), \qquad (13)
\end{aligned}$$

where we used that $h_t^2 \le h_t$.

Denote by

$$\widetilde{\boldsymbol{c}} = \operatorname*{arg\,min}_{\boldsymbol{c}\in\mathcal{C}} \left\{ \exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j))^2\Big) + \lambda \sum_{j=1}^{K} c_j \right\}$$
$$\boldsymbol{c}^\star = \operatorname*{arg\,min}_{\boldsymbol{c}\in[0,1]^K} \left\{ \exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j))^2\Big) + \lambda \sum_{j=1}^{K} c_j \right\}$$

Now, since $\mathcal{C}$ is such that for any $c^\star \in [0,1]$ there is a $\widetilde{c} \in \mathcal{C}$ that satisfies $|\widetilde{c} - c^\star| \le \varepsilon$ we have that

$$\begin{aligned}
&\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(\widetilde{c}_j))^2\Big) + \lambda\sum_{j=1}^{K}\widetilde{c}_j - \Big(\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j^\star))^2\Big) + \lambda\sum_{j=1}^{K} c_j^\star\Big)\\
&\le 4\sum_{j=1}^{K}|p_j(\widetilde{c}_j) - p_j(c_j^\star)| + \lambda\sum_{j=1}^{K}|\widetilde{c}_j - c_j^\star|\\
&\le (4L+\lambda)\sum_{j=1}^{K}|\widetilde{c}_j - c_j^\star| \le (4L+\lambda)K\varepsilon, \qquad (14)
\end{aligned}$$

where in the first inequality we used the same bound as we used in equation (12). By combining equations (13) and (14) we can see that

$$\begin{aligned}
\frac{7}{8}\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_{t,j}))^2\Big) + \lambda\sum_{j=1}^{K} c_j \le \min_{\boldsymbol{c}\in[0,1]^K}&\left\{ \exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j))^2\Big) + \lambda\sum_{j=1}^{K} c_j \right\}\\
&+ (4L+\lambda)K\varepsilon + 96K\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right)
\end{aligned}$$

□

Lemma 4.6. *For any round $t > N$ and $\delta \in (0,1)$, assuming $\mathcal{E}_t^c$ and $\Lambda_{\delta,t}$ both hold,*

$$\begin{aligned}
&\mathbb{P}_{Z_t}\big(\widehat{y}_t \ne y_t \mid \mathcal{F}_t\big) + \lambda\sum_{j=1}^{K} c_{t,j}\\
&\le \textsc{opt}(\boldsymbol{p}) + \frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + (4L+\lambda)K\varepsilon\\
&\quad + 96K\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right).
\end{aligned}$$

Proof. We work in the case where $\widehat{p}_{t,j_t}(c_{t,j_t}) \le \alpha_{t,j_t}(c_{t,j_t})$. The case where $1 - \widehat{p}_{t,j_t}(c_{t,j_t}) \le \alpha_{t,j_t}(c_{t,j_t})$ follows from

symmetric arguments. Since we still work in the case where $\Lambda_{\delta,t}$ holds we have that

$$\begin{aligned}
&\mathbb{E}_Z[\mathbb{1}[\widehat{y}_t \neq y_t]] + \lambda \sum_{j=1}^{K} c_{t,j} = p_{j_t}(c_{t,j_t}) + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\leq \widehat{p}_{t,j_t}(c_{t,j_t}) + |p_{j_t}(c_{t,j_t}) - \widehat{p}_{t,j_t}(c_{t,j_t})| + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\leq \alpha_{t,j_t}(c_{t,j_t}) + \sqrt{\frac{\widehat{p}_{t,j_t}(c_{t,j_t})}{n_{t,j}(c_{t,j_t})} 2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\leq \alpha_{t,j_t}(c_{t,j_t}) + \sqrt{\frac{\alpha_{t,j_t}(c_{t,j_t})}{n_{t,j}(c_{t,j_t})} 2\ln(3/\delta)} + \frac{3\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + \lambda \sum_{j=1}^{K} c_{t,j} \\
&= \frac{\beta + 3\ln(3/\delta) + \sqrt{2\beta\ln(3/\delta)}}{n_{t,j}(c_{t,j_t})} + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\leq \frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + \lambda \sum_{j=1}^{K} c_{t,j} \\
&\leq \min_{\boldsymbol{c}\in[0,1]^K} \left\{ \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_j))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \right\} \\
&\quad + \frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + (4L+\lambda)K\varepsilon + 96K \sum_{j=1}^{K} \left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2 \right),
\end{aligned}$$

where the second inequality follows from the assumption that $\Lambda_{\delta,t}$ holds, the third inequality follows from $\widehat{p}_{t,j_t}(c_{t,j_t}) \leq \alpha_{t,j_t}(c_{t,j_t})$, and the final inequality follows from adding $\frac{7}{8}\Phi(\boldsymbol{p}(\boldsymbol{c}_t))$ and Lemma 4.4. □

Theorem A.2. *Let $\mathcal{C}$ be such that for any $c^\star \in [0,1]$ there is a $\widetilde{c} \in \mathcal{C}$ that satisfies $|\widetilde{c} - c^\star| \leq \varepsilon$ and let $\beta = 18\ln(3/\delta)K^2$. Then for any $\delta \in (0,1)$*

$$\begin{aligned}
&\sum_{t=1}^{T} \mathbb{E}_Z\Big[\mathbb{E}_{\widehat{y}_t}[\mathbb{1}[\widehat{y}_t \neq y_t]] + \lambda \sum_{j=1}^{K} c_{t,j}\Big] \\
&\leq T \min_{\boldsymbol{c}\in[0,1]^K} \left\{ \exp\Big(-2\sum_{j=1}^{K} (\tfrac{1}{2} - p_j(c_j))^2 \Big) + \lambda \sum_{j=1}^{K} c_j \right\} + (1+\lambda K)T^3 K\delta + T(4L+\lambda)K\varepsilon \\
&\quad + N\Big(1 + \lambda K + 2592K^2\big(\ln(3/\delta)\big)^2 + (1+\ln(T))\Big(K(2\beta + 4\ln(3/\delta)) + 576K^2\ln(3/\delta)\Big)\Big)
\end{aligned}$$

Proof. First, since $\mathbb{1}[\widehat{y}_t \neq y_t] + \lambda\sum_{j=1}^{K} c_{t,j} \leq 1 + K\lambda$ we have that

$$\mathbb{E}\left[\sum_{t=1}^{T} \mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\right] \leq N(1+K\lambda) + \mathbb{E}\left[\sum_{t=N+1}^{T} \mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\right].$$

By the Tower rule we have that

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=N+1}^{T}\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\right] = \mathbb{E}\left[\sum_{t=N+1}^{T}\mathbb{E}\left[\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\middle|\mathcal{F}_t\right]\right] \\
&= \mathbb{E}\left[\sum_{t=N+1}^{T}\mathbb{1}[\Lambda_{\delta,t}]\mathbb{E}\left[\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\middle|\mathcal{F}_t\right]\right] + \mathbb{E}\left[\sum_{t=N+1}^{T}\mathbb{1}[\Lambda_{\delta,t}^c]\mathbb{E}\left[\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\middle|\mathcal{F}_t\right]\right] \\
&\leq \mathbb{E}\left[\sum_{t=N+1}^{T}\mathbb{1}[\Lambda_{\delta,t}]\mathbb{E}\left[\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j}\middle|\mathcal{F}_t\right]\right] + \delta K N T^2(\lambda K + 1),
\end{aligned}$$

where the last inequality follows from Lemma 4.1. The remainder of the proof consists of controlling the conditional expectation on the r.h.s. of the above equation. In the remainder of the proof we assume $\Lambda_{\delta,t}$. By Lemmas 4.3, 4.4, 4.5, and 4.6 we have that

$$\begin{aligned}
\mathbb{E}[\mathbb{1}[\widehat{y}_t \neq y_t] + \sum_{j=1}^{K} c_{t,j} \mid \mathcal{F}_t] \leq \min_{\boldsymbol{c}\in[0,1]^K}&\left\{\Phi(\boldsymbol{p}(\boldsymbol{c})) + \lambda\sum_{j=1}^{K} c_j\right\} \\
&+ \frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} + (4L+\lambda)K\varepsilon + 288K\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right)
\end{aligned}$$

For $t \geq N = |\mathcal{C}|$ we have that $n_{t,j}(c) \geq 1$. We continue by bounding

$$\begin{aligned}
&\sum_{t=N+1}^{T}\sum_{j=1}^{K}\left(\frac{2\ln(3/\delta)}{n_{t,j}(c_{t,j})} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c_{t,j})}\Big)^2\right) \\
&= \sum_{t=N+1}^{T}\sum_{j=1}^{K}\sum_{c\in\mathcal{C}}\mathbb{1}[c_{t,j}=c]\left(\frac{2\ln(3/\delta)}{n_{t,j}(c)} + \Big(\frac{3\ln(3/\delta)}{n_{t,j}(c)}\Big)^2\right) \\
&\leq KN\big((1+\ln(T))2\ln(3/\delta) + 9\big(\ln(3/\delta)\big)^2\big),
\end{aligned}$$

where the inequality is $\sum_{t=1}^{T}\frac{1}{t} \leq 1+\ln(T)$ and $\sum_{t=1}^{T}\frac{1}{t^2} \leq 2$. Similarly, we have

$$\sum_{t=N+1}^{T}\frac{2\beta + 4\ln(3/\delta)}{n_{t,j_t}(c_{t,j_t})} \leq KN(2\beta + 4\ln(3/\delta))(1+\ln(T))$$

Thus, combining the above equations we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{1}[\widehat{y}_t \neq y_t]] + \lambda\sum_{j=1}^{K} c_{t,j}\right] \\
&\leq T\min_{\boldsymbol{c}\in[0,1]^K}\left\{\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j))^2\Big) + \lambda\sum_{j=1}^{K} c_j\right\} + N(1+\lambda K) + (1+\lambda K)NT^2K\delta \\
&\quad + KN(2\beta + 4\ln(3/\delta))(1+\ln(T)) + T(4L+\lambda)K\varepsilon + 288K^2N\big((1+\ln(T))2\ln(3/\delta) + 9\big(\ln(3/\delta)\big)^2\big) \\
&= T\min_{\boldsymbol{c}\in[0,1]^K}\left\{\exp\Big(-2\sum_{j=1}^{K}(\tfrac{1}{2} - p_j(c_j))^2\Big) + \lambda\sum_{j=1}^{K} c_j\right\} + (1+\lambda K)NT^2K\delta + T(4L+\lambda)K\varepsilon \\
&\quad + N\Big(1 + \lambda K + 2592K^2\big(\ln(3/\delta)\big)^2 + (1+\ln(T))\Big(K(2\beta + 4\ln(3/\delta)) + 576K^2\ln(3/\delta)\Big)\Big),
\end{aligned}$$

which completes the proof after replacing $\beta = 8\ln(3/\delta)K^2$. □

B. ADDITIONAL EXPERIMENTS

Here we present the additional experimental results announced in Section 6.

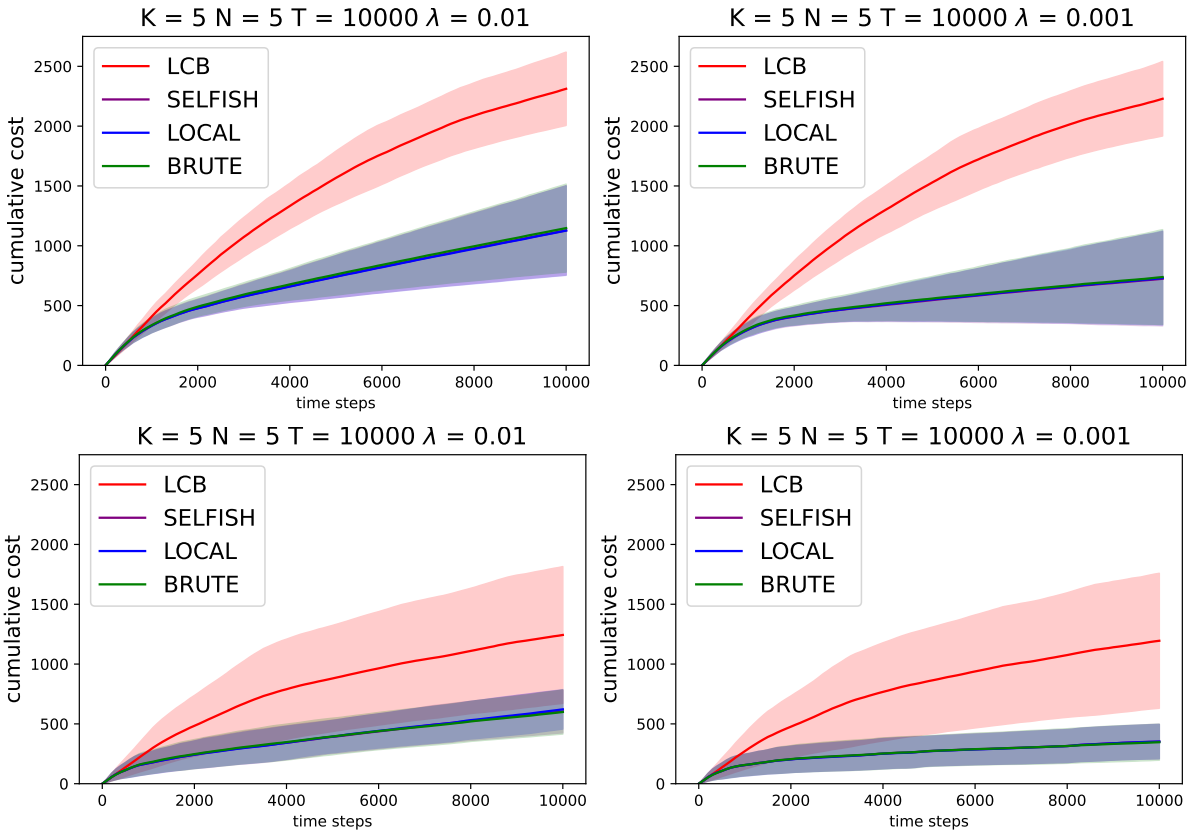

Figure 3. The top two figures and the bottom two figures show the performance of algorithms on the first and second productivity function, respectively.

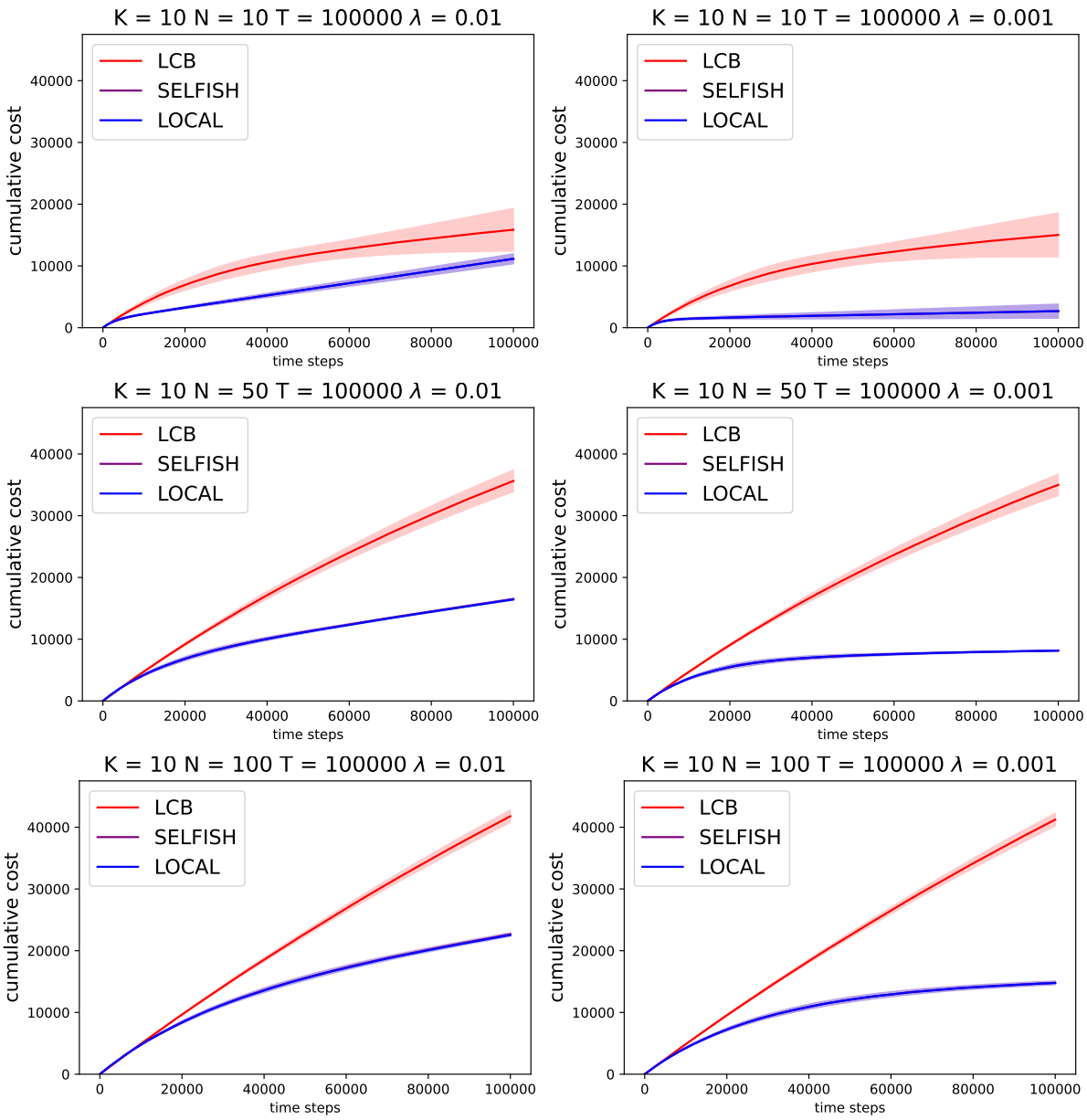

Figure 4. Plots showing cumulative cost when the algorithms are using the first productivity function for $T = 10^5$, $K = 10$, $N \in \{10, 50, 100\}$, and $\lambda \in \{10^{-2}, 10^{-3}\}$.

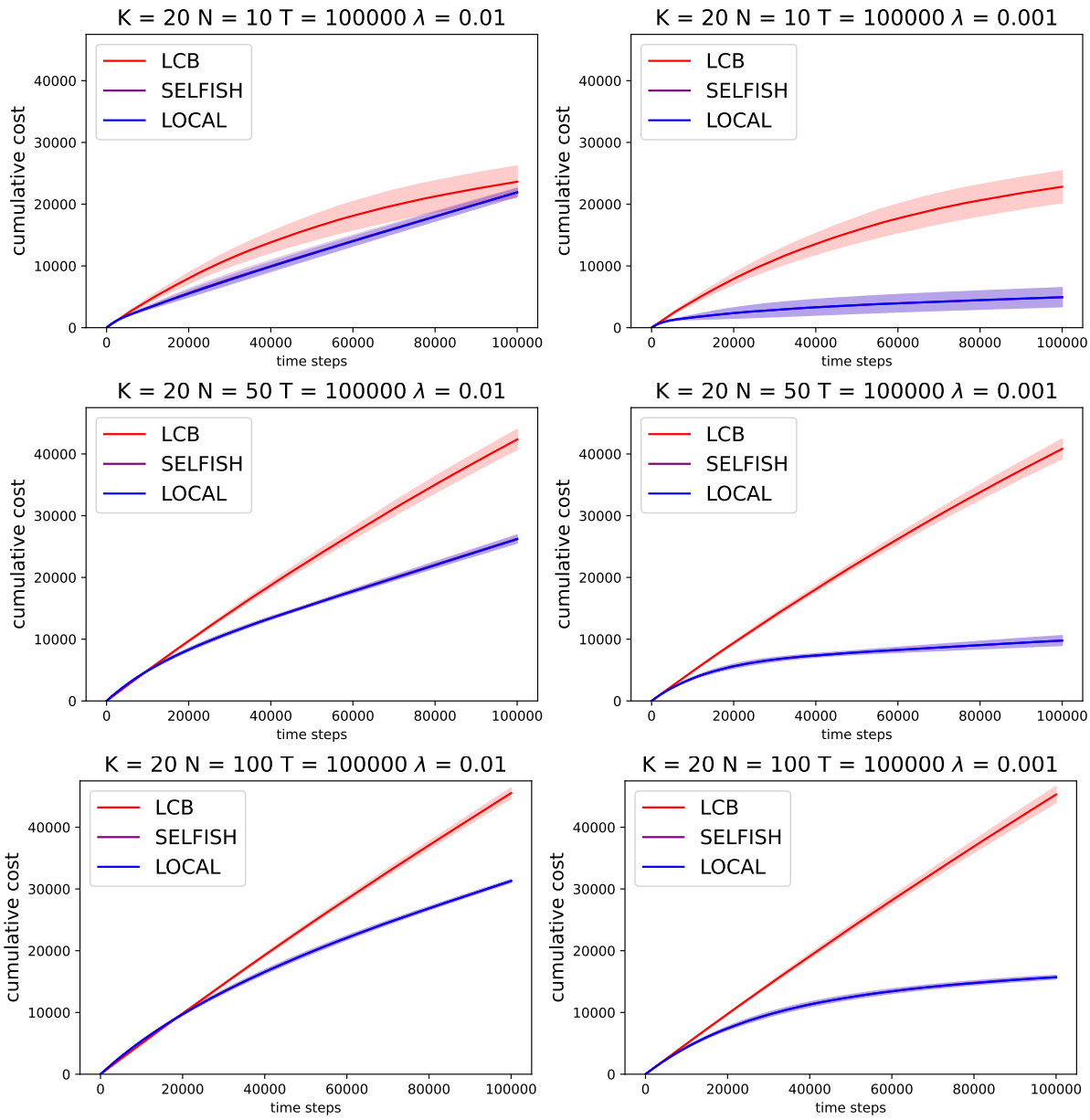

Figure 5. Plots showing cumulative cost when the algorithms are using the first productivity function for $T = 10^5$, $K = 20$, $N \in \{10, 50, 100\}$, and $\lambda \in \{10^{-2}, 10^{-3}\}$.

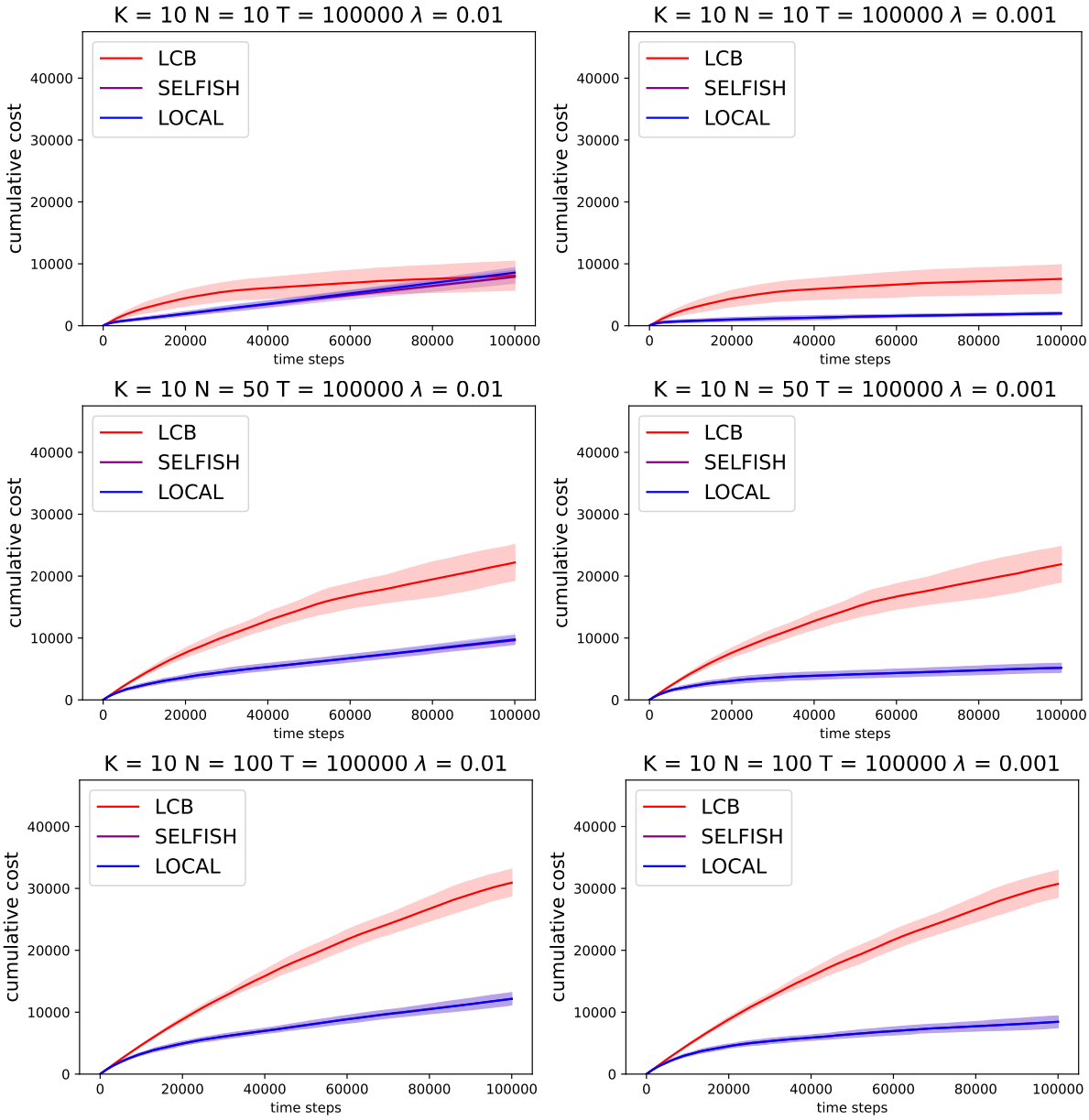

Figure 6. Plots showing cumulative cost when the algorithms are using the second productivity function for $T = 10^5$, $K = 10$, $N \in \{10, 50, 100\}$, and $\lambda \in \{10^{-2}, 10^{-3}\}$.

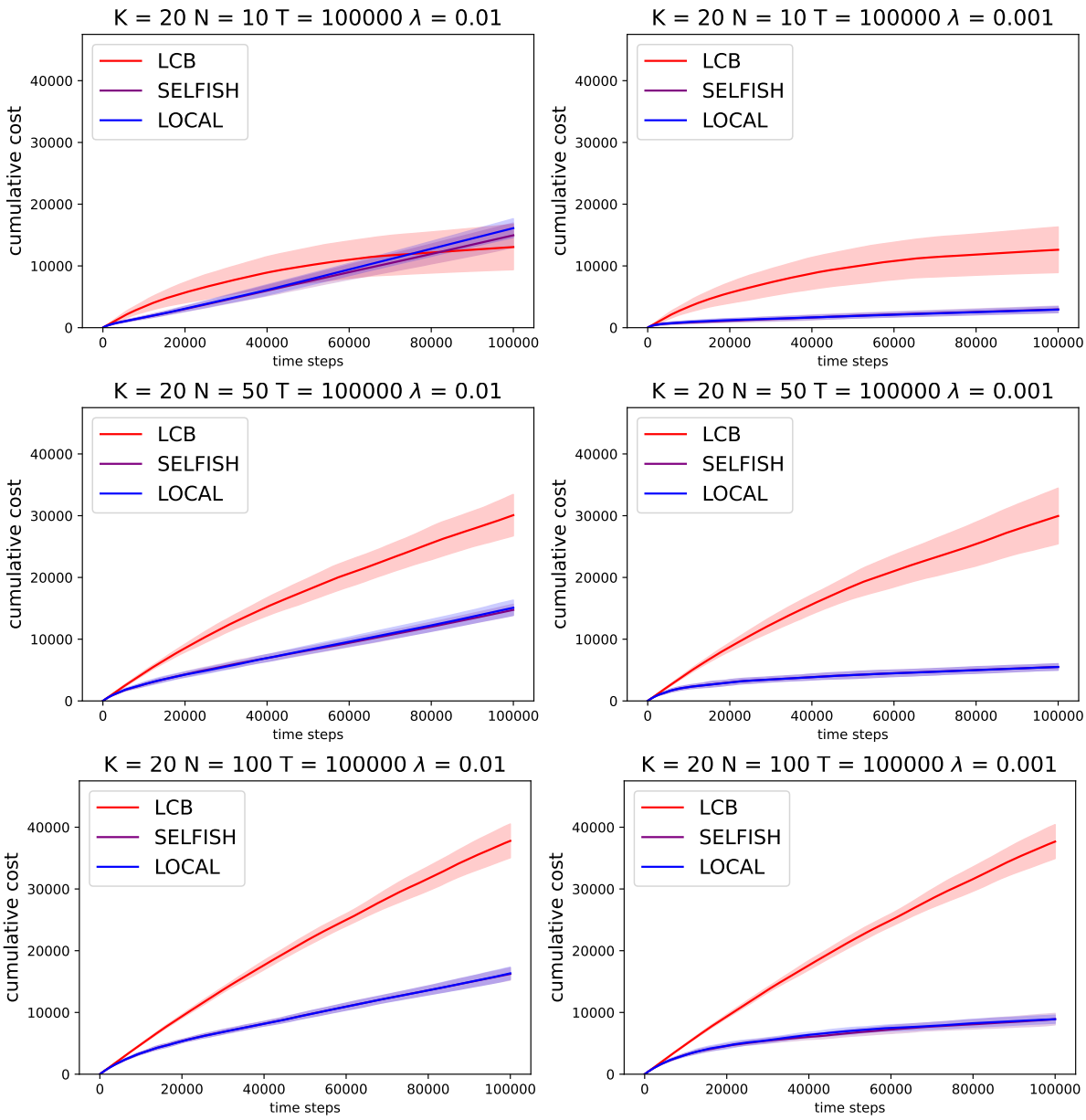

Figure 7. Plots showing cumulative cost when the algorithms are using the second productivity function for $T = 10^5$, $K = 20$, $N \in \{10, 50, 100\}$, and $\lambda \in \{10^{-2}, 10^{-3}\}$.