Adaptive Data Debiasing Through Bounded Exploration

Yifan Yang [1] **Yang Liu** [2] **Parinaz Naghizadeh** [1]

Abstract

Biases in existing datasets used to train algorithmic decision rules can raise ethical and economic concerns due to the resulting disparate treatment of different groups. We propose an algorithm for sequentially debiasing such datasets through adaptive and bounded exploration in a classification problem with costly and censored feedback. Our proposed algorithm includes parameters that can be used to balance between the ultimate goal of removing data biases – which will in turn lead to more accurate and fair decisions, and the exploration risks incurred to achieve this goal. We analytically show that such exploration can help debias data in certain distributions. We further investigate how fairness criteria can work in conjunction with our data debiasing algorithm. We illustrate the performance of our algorithm using experiments on synthetic and real-world datasets.

1. Introduction

Data-driven algorithmic decision making is being adopted widely to aid humans' decisions, in applications ranging from loan approvals to determining recidivism in courts. However, the datasets used for training these algorithms might not accurately represent the agents they make decisions on, due to, e.g., historical biases in decision making and feature selection, or changes in the populations' characteristics or participation since the data was initially collected. This in turn can result in disparate treatment of underrepresented or disadvantaged groups. Motivated by this, we propose an algorithm which, while attempting to make accurate (and fair) decisions, also aims to recover unbiased estimates of the characteristics of agents interacting with it.

In particular, we study a classification problems with *censored and costly feedback*. Censored feedback means that the decision maker only observes the true qualification state of those individuals it admits (e.g., a bank will only observe whether an individual defaults on or repays a loan if the loan is extended in the first place). In such settings, any mismatch between the available training data and the true population may grow over time due to adaptive sampling bias: once a decision rule is adopted based on the current training data, the algorithm's decisions will impact new data collected in the future, in that only agents passing the requirements set by the current decision rule will be admitted going forward. In response, the decision maker may attempt to collect more data from the population; however, such data collection is costly (e.g., may require extending loans to unqualified individuals). Given these challenges, we present an *active debiasing* algorithm with *bounded exploration*: our algorithm admits some agents that would otherwise be rejected (i.e., it explores), yet adaptively and judiciously limits the extent and frequency of this exploration.

In particular, in each time period, our algorithm selects a (fairness-constrained) decision rule that minimizes classification error based on its current, possibly biased training data; adopting this decision rule corresponds to *exploitation* of the current information by the algorithm. At the same time, to circumvent the censored feedback nature of the problem, our algorithm also deviates from the prescriptions of this loss-minimizing classifier to a judiciously chosen extent (the extent is chosen adaptively, based on the current estimates); this will constitute *exploration*. Our algorithm includes two parameters to limit the costs of this exploration: one modulates the *frequency* of exploration (an exploration probability ϵ_t which can be adjusted using current bias estimates), and another limits the *depth* of exploration (by setting a threshold LB_t on how far from the classifier one is willing to go when exploring).

Summary of findings and contributions.

1. *Comparison with baselines.* We contrast our proposed algorithm against two baselines: `exploitation-only` (one that does not include any form of exploration), and `pure exploration` (which may randomly accept some of the agents rejected by the classifier, but does not bound exploration). We show (Theorem 4.1) that `exploitation-only` always overestimates of the underlying distributions. Also, while `pure exploration`

[1]Department of Integrated Systems Engineering, The Ohio State University, Columbus, Ohio, USA [2]Department of Computer Science and Engineering, University of California Santa Cruz, Santa Cruz, California, USA. Correspondence to: Yifan Yang <yang.5483@osu.edu>.

ICML 2022 Workshop on Responsible Decision Making in Dynamic Environments, Baltimore, Maryland, USA, 2022.

can debias the estimates in the long-run (Theorem 4.2), it does so at the expense of accepting *any* agent.

2. *Analytical support for our proposed algorithm.* We show (Theorem 4.3) that our proposed `active debiasing` algorithm with bounded exploration can correct biases in unimodel distribution estimates. We also provide an error bound for our algorithm (Theorem 4.4).

3. *Interplay with fairness criteria.* We analyze the impact of fairness constraints on our algorithm's performance, and show (Proposition 4.5) that existing fairness criteria may speed up debiasing of the data in one group, while slowing it down for another.

4. *Numerical experiments.* We provide numerical support for the performance of our algorithm using experiments on synthetic and real-world (Adult and FICO) datasets.

Related work. Our paper is closely related to the works of (Bechavod et al., 2019; Kilbertus et al., 2020; Ensign et al., 2018; Blum & Stangl, 2020; Jiang & Nachum, 2020), which study the impact of data biases on (fair) algorithmic decision making. Among these works, Bechavod et al. (2019) and Kilbertus et al. (2020) study *fairness-constrained* learning in the presence of censored feedback. While these works also use exploration, the form and purpose of exploration is different: the algorithm in (Bechavod et al., 2019) starts with a pure exploration phase, and subsequently explores with the goal of ensuring the fairness constraint is not violated; the stochastic (or exploring) policies in (Kilbertus et al., 2020) conduct (pure) exploration to address the censored feedback issue. In contrast, we start with a biased dataset, and conduct *bounded* exploration to debias data; fairness constraints may or may not be enforced separately and are orthogonal to our debiasing process.

Our work is also closely related to (Deshpande et al., 2018; Nie et al., 2018; Neel & Roth, 2018; Wei, 2021), which study adaptive sampling biases induced by a decision rule, particularly when feedback is censored. Among these, the recent work of Wei (2021) is most closely related to ours, and studies data collection by formulating the problem as a partially observable Markov decision processes. Using dynamic programming methods, the data collection policy is shown to be a threshold policy that becomes more stringent (in our terminology, reduces exploration) as learning progresses. Our works differ in the problem setup and our analysis of the impact of fairness constraints. More importantly, in contrast to all these works, our starting point is a *biased* dataset (which may be biased for reasons other than adaptive sampling in its collection, including historical biases); we then show how, while attempting to debias this dataset by collecting new data, any additional adaptive sampling bias during data collection can be prevented.

Our work is also broadly related to Bandit learning; additional and detailed discussions are given in Appendix A.

2. Model and Preliminaries

The environment. We consider a *firm* or decision maker, who selects an algorithm to make decisions on a population of *agents*. The firm observes agents arriving over times $t = 1, 2, \ldots$, makes a decision for agents arriving at time t based on the current algorithm, and subsequently adjusts its algorithm for times $t + 1$ based on the observed outcomes.

Each agent has an observable *feature* or *score* $x \in \mathcal{X} \subseteq \mathbb{R}$ representing their characteristics (e.g., credit scores or exam scores). We use a one-dimensional feature setting in our analysis, and generalize to $\mathcal{X} \subseteq \mathbb{R}^n$ in Section 5. They are either qualified or unqualified to receive a favorable decision captured by their true *label* $y \in \{0, 1\}$. In addition, each agent belongs to a *group* based on its protected attributes (e.g., race, gender) denoted as $g \in \{a, b\}$. We consider threshold-based, group-specific, binary classifiers $h_{\theta_{g,t}}(x) = \mathbb{1}(x \geq \theta_{g,t})$ as (part of) the algorithm adopted by the firm, where $\theta_{g,t}$ denotes the classifier's decision threshold. An agent from group g with feature x arriving at time t is admitted iff $x \geq \theta_{g,t}$.

Quantifying bias. Let $f_g^y(x) = \mathbb{P}(X = x | Y = y, G = g)$ denote the true underlying pdf for the feature distribution of agents from group g with label y. The algorithm has an estimate of these unknown distributions, at each time t, based on the data collected so far (or an initial training set). Denote the algorithm's estimate at t by $\hat{f}_{g,t}^y(x)$.

Assumption 2.1. The firm updates its estimates $\hat{f}_{g,t}^y(x)$ by updating a single parameter $\hat{\omega}_{g,t}^y$.

Under Assumption 2.1, the bias can be captured by the mistmatch between the estimated and true parameters $\hat{\omega}_{g,t}^y$ and ω_g^y. In particular, we set the mean absolute error $\mathbb{E}[|\hat{\omega}_{g,t}^y - \omega_g^y|]$ as the measure for quantifying bias.

Algorithm choice without debiasing. Let α_g^y be the fraction of group g agents with label y. A loss-minimizing fair algorithm selects its thresholds $\theta_{g,t}$ at time t as follows:

$$\min_{\theta_{a,t}, \theta_{b,t}} \sum_{g \in \{a,b\}} \alpha_g^1 \int_{-\infty}^{\theta_{g,t}} \hat{f}_{g,t}^1 \mathrm{d}x + \alpha_g^0 \int_{\theta_{g,t}}^{\infty} \hat{f}_{g,t}^0 \mathrm{d}x$$
$$\text{s.t.} \quad \mathcal{C}(\theta_{a,t}, \theta_{b,t}) = 0 \,. \tag{1}$$

Here, the objective is the misclassification error, and $\mathcal{C}(\theta_a, \theta_b) = 0$ is the fairness constraint imposed by the firm, if any. For instance, $\mathcal{C}(\theta_{a,t}, \theta_{b,t}) = \theta_{a,t} - \theta_{b,t}$ for *same decision rule*, or $\mathcal{C}(\theta_{a,t}, \theta_{b,t}) = \int_{\theta_{a,t}}^{\infty} \hat{f}_{a,t}^1(x)\mathrm{d}x - \int_{\theta_{b,t}}^{\infty} \hat{f}_{b,t}^1(x)\mathrm{d}x$ for *equality of opportunity*.

3. An `Active Debiasing` Algorithm with Bounded Exploration

In this section, we present the `active debiasing` algorithm which uses both *exploitation* (the decision rules

of (1)) and *exploration* (some deviations) to remove any biases from the estimates $\hat{f}^y_{g,t}$. Although the deviations may lead to admission of some unqualified agents, they can be beneficial to the firm in the long-run: by reducing biases in $\hat{f}^y_{g,t}$, both classification loss estimates and fairness constraint evaluations can be improved. In this section, we drop the subscripts g from the notation; when there are multiple groups, our algorithm is applied to each group separately.

As noted in Section 1, our algorithm is one of *bounded* exploration: it includes a *lowerbound* LB_t, which captures the extent to which the decision maker is willing to deviate from the current classifier θ_t, based on its current estimate $\hat{f}^0_t$ of the unqualified agents' underlying distribution. Formally,

Definition 3.1. At time t, the firm selects a LB_t such that

$$\text{LB}_t = (\hat{F}^0_t)^{-1}(2\hat{F}^0_t(\hat{\omega}^0_t) - \hat{F}^0_t(\theta_t))$$

where θ_t is the (current) threshold determined from (1), $\hat{F}^0_t$, $(\hat{F}^0_t)^{-1}$ are the cdf and inverse cdf of the estimates $\hat{f}^0_t$, respectively, and $\hat{\omega}^0_t$ is (wlog) the α-th percentile of $\hat{f}^0_t$.

Notice that by selecting a high α-th percentile in the above definition, LB_t can be increased so as to limit the depth of exploration. As shown later, this thresholding choice will enable debiasing while controlling its costs.

Algorithm 1 (The `active debiasing` algorithm). Let the decision threshold be θ_t, and LB_t be given by Definition 3.1. Let $\{\epsilon_t\}$ be a sequence of exploration probabilities. For agents $(x^\dagger, y^\dagger)$ arriving at time t:

Step I: Admit agents and collect data. Admit all agents with $x^\dagger \geq \theta_t$. Additionally, if $\text{LB}_t \leq x^\dagger < \theta_t$, admit the agent with probability ϵ_t.

Step II: Update the distribution estimates based on new data collected in Step I. Identify new data with $\text{LB}_t \leq x^\dagger$ and $y^\dagger = 1$ (resp. $y^\dagger = 0$). Use all such $x^\dagger$ with $\text{LB}_t \leq x^\dagger < \theta_t$, and such $x^\dagger$ with $\theta_t \leq x^\dagger$ with probability ϵ_t, to update $\hat{\omega}^1_t$ (resp. $\hat{\omega}^0_t$).

4. Theoretical Analysis

We consider two baselines: `exploitation-only` and `pure exploration`, and highlight the benefits of bounded exploration through our `active debiasing` algorithm. All proofs are given in Appendix B.

4.1. The `exploitation-only` baseline

Our first baseline algorithm only updates its estimates of the underlying distributions based on agents with $x \geq \theta_t$ who pass the (current) loss-minimizing classifier (1). The following result shows that this approach consistently suffers from adaptive sampling bias, ultimately resulting in overestimation of the underlying distributions.

Theorem 4.1. *An* `exploitation-only` *algorithm overestimates ω^y, i.e., $\lim_{t\to\infty} \mathbb{E}[\hat{\omega}^y_t] > \omega^y, \forall y$.*

4.2. The `pure exploration` baseline

In this second baseline, at each time t, the algorithm may accept any agent with $x < \theta_t$ with probability ϵ_t. The following result establishes that using the data collected this way, the distributions can be debiased in the long-run, *if* the data collected above the classifier is also sampled with probability ϵ_t when updating the distributions.

Theorem 4.2. *Using the* `pure exploration` *algorithm, $\hat{\omega}^y_t \to \omega^y$ as $t \to \infty$, $\forall y$.*

4.3. The `active debiasing` algorithm

While `pure exploration` can successfully debias data in the long-run, it does so at the expense of accepting agents with *any* $x < \theta_t$. The following result shows that our algorithm can still debias data in certain distributions, while limiting the depth of exploration to $\text{LB}_t < x < \theta_t$.

Theorem 4.3. *Let f^y and $\hat{f}^y_t$ denote the true feature distribution and their estimates at the beginning of time t, with respective α-th percentiles ω^y and $\hat{\omega}^y_t$. Assume these are unimodel distributions, $\epsilon_t > 0, \forall t$, and $\hat{\omega}^0_t \leq \theta_t \leq \hat{\omega}^1_t, \forall t$. Then, using the* `active debiasing` *algorithm,*
***(a)** If $\hat{\omega}^y_t$ is underestimated (resp. overestimated), then $\mathbb{E}[\hat{\omega}^y_{t+1}] \geq \hat{\omega}^y_t$, (resp. $\mathbb{E}[\hat{\omega}^y_{t+1}] \leq \hat{\omega}^y_t$) $\forall t, \forall y$.*
***(b)** The sequence $\{\hat{\omega}^y_t\}$ converges to ω^y as $t \to \infty$, $\forall y$.*

4.4. Error bound analysis

Our error bound analysis compares the errors (measured as the number of wrong decisions made) of our `adaptive debiasing` algorithm with the errors made by an oracle which knows the true underlying distributions. The following theorem provides an upperbound on the errors incurred by `active debiasing`.

Theorem 4.4. *Let $\hat{f}^y_{g,t}(x)$ be the estimated feature distributions at round $t \in \{1, \ldots, m\}$, and $\theta_{g,t}$ be a v-approximate solution. Denote the Rademacher complexity of the classifier family $\mathcal{H}$ with n training samples by $\mathcal{R}_n(\mathcal{H})$. At round t, among the two groups, let N_t be the larger net exploration errors made, n'_t be the smaller sample size collected, $\mathcal{R}_{n''_t}(\mathcal{H})$ be the larger Rademacher complexity. Then, with probability at least $1 - 4\delta$ with $\delta > 0$, the active debiasing algorithm has an error bound (Err.):*

$$\textit{Err.} \leq 8m\mathcal{R}_{n''_1}(\mathcal{H}) + \frac{8m}{\sqrt{n'_1}} + 2m\sqrt{\frac{2\ln(2/\delta)}{n'_1}} + 2mN_1 + 4mv$$

4.5. `Active debiasing` and fairness criteria

We consider our proposed algorithm when used in conjunction with fairness constraints (e.g., equality of opportunity, same decision rule). Imposing such fairness rules will lead to changes to the selected classifiers compared to the fairness-unconstrained case. Let $\theta^F_{g,t}$ and $\theta^U_{g,t}$ be the fairness constrained and unconstrained decision rules obtained from (1). We say group g is being over-selected (resp. under-selected) if $\theta^F_{g,t} < \theta^U_{g,t}$ (resp. $\theta^F_{g,t} > \theta^U_{g,t}$). The following

result shows how such over/under-selections can differently affect the debiasing of estimates on different agents.

Proposition 4.5. *Let f_g^y and $\hat{f}_{g,t}^y$ be the true and estimated feature distributions, with respective medians ω^y and $\hat{\omega}_t^y$. Assume these are unimodel distributions, and* `active debiasing` *is applied. If group g is over-selected (resp. under-selected) under a fairness constraint, i.e., $\theta_{g,t}^F < \theta_{g,t}^U$ (resp. $\theta_{g,t}^F > \theta_{g,t}^U$), the speed of debiasing on the estimates $\hat{f}_{g,t}^y$ will decrease (resp. increase).*

5. Numerical Experiments

In this section, we illustrate the performance of our algorithm through numerical experiments on both Gaussian and Beta distributed synthetic datasets, and on two real-world datasets: the *Adult* dataset (Dua & Graff, 2017) and the *FICO* credit score dataset (Reserve, 2007) pre-processed by (Hardt et al., 2016). More figures are in Appendix C.

Performance on Beta distributions: Fig. 1(a) shows that our algorithm can debias data for Beta distributions with a mistmach between the α parameter. This verifies that Theorem 4.3 can hold beyond symmetric distributions.

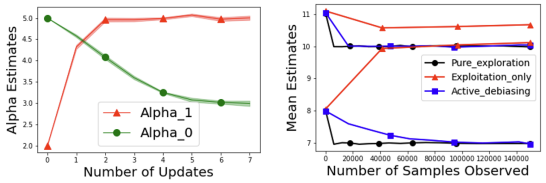


Figure 1. Debiasing on Beta and comparison with baselines.

Comparison with baselines: Our first Gaussian experiments in Fig. 1(b), compare our algorithm against two baselines. It shows that consistent with Theorem 4.1, `exploitation-only` overestimates the distributions due to adaptive sampling biases. We also observe that as expected, `pure exploration` debiases *faster* than `active debiasing`, but as shown next, incurs higher exploration costs while doing so.

Regret and Weighted Regret: Figs. 2 compare the regret and weighted regret of the algorithms. Regret is measured as the difference between the number of wrong decisions made by our algorithm vs the oracle classifier. Weighted regret is defined similarly, but also adds an exponential weight to each wrong decision. We observe that `exploitation-only`'s regret is super-linear, as not only it fails to debias, but has increasing error due to biases from overestimating. On the other hand, while algorithms that explore "deeper" have lower regret in Fig. 2(a), they have higher weighted regret shown in Fig. 2(b).

Interplay of debiasing and fairness constraints: Fig. 3 compare the performance when there are two groups of agents with underlying Gaussian distributions, and the algorithm is chosen subject to three different fairness settings: no fairness, equality of opportunity (EO), and the same decision rule (SD). The findings are consistent with Prop. 4.5. For instance, SD will over-select the advantaged group so that, as shown in the left panel in Fig. 3, the speed of debiasing on the estimates $\hat{f}_{a,t}^y$ will decrease. In contrast, an opposite effect will happen in the disadvantaged group b.

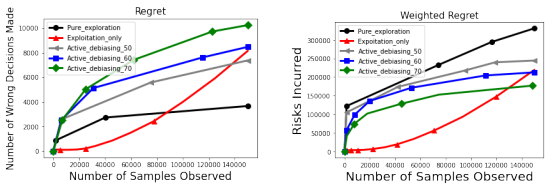


Figure 2. Regret and weighted regret.

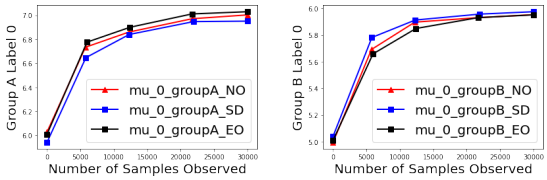
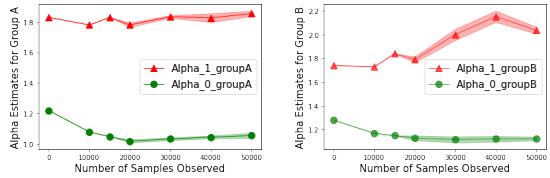
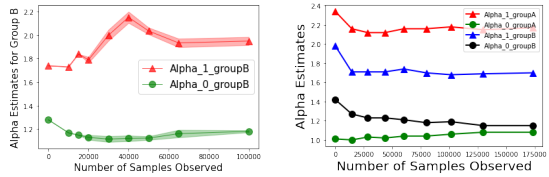


Figure 3. `Debiasing` used with fairness constraints.

`Active debiasing` on the *Adult* and *FICO* dataset: Fig. 4 (a-c) and (d) illustrate the performance on the *Adult* and *FICO* dataset. We observe that our proposed algorithm



(a) Debiasing G_a, Adult. (b) Debiasing G_b, Adult.



(c) G_b with additional data. (d) `Debiasing` on *FICO*.

Figure 4. `Active debiasing` on the *Adult* and *FICO* datasets.

can debias estimates across groups and for both labels, but that this happens in the long-run with sufficient samples: in *Adult*, as there are only 1080 samples for label 1 agents from G_b, although the bias initially decreases, the final estimate still differs from the true value. Fig. 4(c) verifies that this estimate would have been debiased in the long-run with additional samples from the underlying population.

Acknowledgements. The authors are grateful for support from Cisco Research, and the NSF program on Fairness in AI in collaboration with Amazon under Award No. IIS-2040800. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF, Amazon, or Cisco.

References

Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., and Wallach, H. A reductions approach to fair classification. In *International Conference on Machine Learning*, pp. 60–69. PMLR, 2018.

Balcan, M.-F., Broder, A., and Zhang, T. Margin based active learning. In *International Conference on Computational Learning Theory*, 2007.

Bechavod, Y., Ligett, K., Roth, A., Waggoner, B., and Wu, S. Z. Equal opportunity in online classification with partial feedback. In *Advances in Neural Information Processing Systems*, pp. 8974–8984, 2019.

Blum, A. and Stangl, K. Recovering from biased data: Can fairness constraints improve accuracy? In *1st Symposium on Foundations of Responsible Computing (FORC 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020.

Corbett-Davies, S., Pierson, E., Feller, A., Goel, S., and Huq, A. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd acm sigkdd international conference on knowledge discovery and data mining*, pp. 797–806, 2017.

Deshpande, Y., Mackey, L., Syrgkanis, V., and Taddy, M. Accurate inference for adaptive linear models. In *International Conference on Machine Learning*, pp. 1194–1203, 2018.

Dua, D. and Graff, C. UCI machine learning repository, 2017. URL http://archive.ics.uci.edu/ml.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. Fairness through awareness. In *Proceedings of the 3rd innovations in theoretical computer science conference*, 2012.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., and Venkatasubramanian, S. Runaway feedback loops in predictive policing. In *Conference on Fairness, Accountability and Transparency*, pp. 160–171. PMLR, 2018.

Hardt, M., Price, E., and Srebro, N. Equality of opportunity in supervised learning. In *Advances in neural information processing systems*, pp. 3315–3323, 2016.

Jiang, H. and Nachum, O. Identifying and correcting label bias in machine learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 702–712, 2020.

Kazerouni, A., Zhao, Q., Xie, J., Tata, S., and Najork, M. Active learning for skewed data sets. *arXiv preprint arXiv:2005.11442*, 2020.

Kilbertus, N., Rodriguez, M. G., Schölkopf, B., Muandet, K., and Valera, I. Fair decisions despite imperfect predictions. In *International Conference on Artificial Intelligence and Statistics*, pp. 277–287. PMLR, 2020.

Liu, L. T., Dean, S., Rolf, E., Simchowitz, M., and Hardt, M. Delayed impact of fair machine learning. In *International Conference on Machine Learning*, pp. 3150–3158. PMLR, 2018.

Liu, L. T., Wilson, A., Haghtalab, N., Kalai, A. T., Borgs, C., and Chayes, J. The disparate equilibria of algorithmic decision making when individuals invest rationally. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.

Maritz, J. and Jarrett, R. A note on estimating the variance of the sample median. *Journal of the American Statistical Association*, 73(361):194–196, 1978.

Neel, S. and Roth, A. Mitigating bias in adaptive data gathering via differential privacy. In *International Conference on Machine Learning*, pp. 3720–3729. PMLR, 2018.

Nie, X., Tian, X., Taylor, J., and Zou, J. Why adaptively collected data have negative bias and how to correct for it. In *International Conference on Artificial Intelligence and Statistics*, pp. 1261–1269, 2018.

Perdomo, J., Zrnic, T., Mendler-Dünner, C., and Hardt, M. Performative prediction. In *International Conference on Machine Learning*, pp. 7599–7609. PMLR, 2020.

Reserve, U. F. Report to the congress on credit scoring and its effects on the availability and affordability of credit. https://www.federalreserve.gov/boarddocs/rptcongress/creditscore/creditscore.pdf, 2007.

Wei, D. Decision-making under selective labels: Optimal finite-domain policies and beyond. In *International Conference on Machine Learning*, pp. 11035–11046. PMLR, 2021.

Zhang, X., Khaliligarekani, M., Tekin, C., and Liu, M. Group retention when using machine learning in sequential decision making: the interplay between user dynamics and fairness. In *Advances in Neural Information Processing Systems*, pp. 15269–15278, 2019.

A. Additional and Detailed Related Work

Data debiasing with censored and costly feedback: Our paper is most closely related to the works of (Ensign et al., 2018; Bechavod et al., 2019; Kilbertus et al., 2020; Blum & Stangl, 2020; Jiang & Nachum, 2020), who have investigated the impacts of data biases on (fair) algorithmic decision making. Ensign et al. (2018)'s work was one of the earliest to identify the feedback loops between predictive algorithms and biases in the collected data; we investigate similar feedback loops, but are primarily focused on debiasing data, as well as the impact of fairness-constrained learning. Bechavod et al. (2019) and Kilbertus et al. (2020) study fairness-constrained learning in the presence of censored feedback. While these works also use exploration, the form and purpose of exploration is different: the algorithm in (Bechavod et al., 2019) starts with a pure exploration phase, and subsequently explores with the goal of ensuring the fairness constraint is not violated; the stochastic (or exploring) policies in (Kilbertus et al., 2020) conduct (pure) exploration to address the censored feedback issue. In contrast, we start with a biased dataset, and conduct *bounded* exploration with the goal of data debiasing while accounting for the costs of exploration; fairness constraints may or may not be enforced separately and are orthogonal to our debiasing process. As shown in Section 5, such pure exploration processes incur higher exploration costs than our proposed bounded exploration algorithm.

A number of other works, including (Deshpande et al., 2018; Nie et al., 2018; Neel & Roth, 2018; Wei, 2021) have, similar to our work, explored the question of biases induced by a decision rule on data collection, particularly when feedback is censored. Deshpande et al. (2018) study inference in a linear model with adaptively collected data; in contrast to our proposed method, their work focuses on debiasing of an estimator, rather than modifying the decision rule used to collect the data. Nie et al. (2018) study the problem of estimating statistical parameters from adaptively collected data. Their proposed adaptive data collection method, which also similar to ours (Assumption 2.1) is used for single-parameter estimation, is one of online debiasing; our proposed data collection methods however differ. In particular, our focus is on accounting for multiple subgroups as well as fairness considerations. More importantly, we propose a *bounded* exploration strategy, which accounts for the risks of exploration decisions and limits the depth of exploration; this method of exploration is different from the random exploration used to collect the data in (Nie et al., 2018), to which their proposed debiasing algorithms based on data splitting and modified maximum likelihood estimators are applied.

While (Deshpande et al., 2018; Nie et al., 2018) propose ex-post methods for debiasing adaptively collected data, Neel & Roth (2018) consider an adaptive data gathering procedure, and show that no debiasing will be necessary if the data is collected through a differentially private method. We similarly propose a debiasing algorithm that adaptively adjusts its data collection procedure, but unlike (Neel & Roth, 2018), account for the costs of exploration in our data collection procedure. The recent work of Wei (2021) studies data collection in the presence of censored feedback, and similar to our work, accounts for the cost of exploration in data collection, by formulating the problem as a partially observable Markov decision processes. Using dynamic programming methods, the data collection policy is shown to be a threshold policy that becomes more stringent (in our terminology, reduces exploration) as learning progresses. Our works are similar in that we both propose using adaptive and cost-sensitive exploration, but we differ in the problem setup and our analysis of the impact of fairness constraints. More importantly, in contrast to both (Neel & Roth, 2018; Wei, 2021), our starting point is a *biased* dataset (which may be biased for reasons other than adaptive sampling in its collection); we then consider how, while attempting to debias this dataset by collecting new data, any additional adaptive sampling bias during data collection should be prevented.

Interplay of fairness criteria and data biases: Our analysis in Section 4.5, similar to those of Blum & Stangl (2020) and Jiang & Nachum (2020), also considers the interplay between algorithmic fairness rules and data biases. Blum & Stangl (2020) show that certain fairness constraints can *themselves* be interpreted as enabling debaising of the underlying estimates. Also, both works study data bias arising due to the *labeling* process and propose reweighting techniques to address it. Our work differs from these from two main aspects. First, we model biases as changes in feature-label distributions, in contrast to the assumption of noisy labels in these works. Second, we introduce a statistical debiasing technique based primarily on exploration, which is orthogonal to the social debiasing achieved through fairness constraints. Our proposed model and algorithm therefore complement these works.

Relation to the bandit learning literature: More broadly, our work is related to the literature on Bandit learning and its study of exploration and exploitation trade-offs, where adaptively adjusted exploration decisions play a key role in allowing the decision maker to attain new information, while at the same time using the collected information to maximize some notion of long-term reward. In particular, bandit exploration deviates from choosing the current best arm in several ways: randomly as in ϵ-greedy, by some form of highest uncertainty as in UCB, by importance sampling approaches as in EXP3,

etc. A key difference of our work with these existing approaches is our choice of *bounded* exploration, where the bounds are motivated by settings in which the cost of wrong decisions increase as samples further away from the current decision threshold are admitted. In that sense, our proposed approach can be viewed as a bounded version of ϵ-greedy; we refer to the non-bounded version of ϵ-greedy in our setting as `pure exploration`, and show that our proposed algorithm can achieve lower weighted regret (one that accounts for the cost of wrong decisions) than `pure exploration`.

Long-term fairness and bias in algorithmic decision making: The majority of works on fair algorithmic decision making have focused on achieving fairness in a one-shot setting (i.e. without regards to the long-term effects of the proposed algorithms); see e.g. (Hardt et al., 2016; Dwork et al., 2012; Corbett-Davies et al., 2017). Some recent works have studied long-term impacts of fairness on disparities, group representation, and strategic manipulation of features, as a result of adopting fairness measures (Liu et al., 2018; Zhang et al., 2019; Liu et al., 2020). Our work contributes to this line of research, by analyzing the long-term effects of imposing fairness constraints on data collection and debiasing efforts.

Relation to the active learning literature: Our work is also related to the active learning literature. Balcan et al. (2007) study the sample complexity of labeled data required for active learning, and Kazerouni et al. (2020) propose an algorithm involving exploration and exploitation-based adaptive sampling, verifying it using simulations. Similar to these works, we use exploration for addressing adaptive sampling bias; the main difference, aside from the application, is in our analytical guarantees as well as our focus on the interplay of debiasing with fairness constraints.

Performative prediction: Finally, the recent line of work on performative prediction proposed by Perdomo et al. (2020) also considers the effects of algorithmic decisions on the underlying population's features-label distributions. In particular, the choice of the ML model can cause a shift in the data distribution, and the goal of this work is to identify the stable ML model parameter that is attained at a fixed point of the algorithm-population interactions. In contrast to this goal, our focus is on *pre-existing* and unchanging distribution shifts in the data, which our `active debiasing` algorithm aims to correct over time. Therefore, our algorithm could be considered as a debiasing method to be used when such performative shifts are present in the data, but are unaccounted for: If distribution shifts happen relatively slower than our debiasing algorithm's convergence speed, our `active debiasing` could be used to recover correct estimates of the underlying distribution, the estimates of which might have been biased due to performative distribution shifts.

B. Proof of Section 4

B.1. Proof of Theorem 4.1

Proof. We detail the proof for label 0 estimates $\hat{\omega}_t^0$, and discuss two cases. First, if $\hat{\omega}_t^0$ is overestimated, i.e. $\hat{\omega}_t^0 > \omega^0$. Note that we have $\theta_t \geq \hat{\omega}_t^0$. Then, as only agents with $x^\dagger \geq \theta_t$ are admitted, $\hat{\omega}_t^0$ may only be updated to stay the same or increase. Therefore, $\hat{\omega}_t^0$ will remain overestimated.

Next consider the case that $\hat{\omega}_t^0$ is underestimated, $\hat{\omega}_t^0 < \omega^0$. From t on, consider the $T \gg t$ next steps. First, since each observation is independently drawn, we know that at time $t' = t, ..., t+T$, $x_{t'} - \mathbb{E}[X|X \geq \theta_{t'}]$ forms a martingale; this is because of the independence of $x_{t'}$ and $\theta_{t'}$ when conditioned on the historical information, as well as the fact that $\mathbb{E}[x_{t'}] = \mathbb{E}[X|X \geq \theta_{t'}]$.

By definition of ω^0, we also know that $\sum_{t'=t}^{T} \mathbb{E}[X|X \geq \theta_{t'}] > T \cdot \omega^0$. Denote the gap by $\Delta := \frac{\sum_{t'=t}^{T} \mathbb{E}[X|X \geq \theta_{t'}]}{T} - \omega^0$. Therefore using the Azuma-Hoeffding inequality we have

$$\mathbb{P}\Big(\sum_{t'=t}^{T} x_{t'} - \sum_{t'=t}^{T} \mathbb{E}[X|X \geq \theta_{t'}] \leq \delta\Big) \leq e^{\frac{-2\delta^2}{T-t+1}},$$

for any $\delta < 0$. Letting $\delta = -\Delta \cdot (T - t + 1)$, the above can be re-written as

$$\mathbb{P}\big(\tfrac{1}{T-t+1} \sum_{t'=t}^{T} x_{t'} > \omega^0\big) > 1 - e^{\left(-2\Delta^2 (T-t+1)\right)} \underbrace{\to 1}_{T \to \infty}$$

This proves that with high probability the mean of the new samples is higher than ω^0. Therefore, at some time T that is significantly higher than t, the new estimate $\hat{\omega}_T^0$ will be similar to $\frac{1}{T-t+1} \sum_{t'=t}^{T} x_{t'}$, which is higher than the true ω^0. From our arguments for the overestimated case, from this point on, $\hat{\omega}_t^0$ will stay overestiamted. The proof for $\hat{\omega}_t^1$ is similar. □

B.2. Proof of Theorem 4.2

The proof follows from assuming (wlog) that the unknown parameter ω^y being estimated is the distribution's mean. Then, as we are collecting i.i.d. samples from across the distribution, $\hat{\omega}_t^y$ can be set to the sample mean, and the conclusion follows from the strong law of large numbers. Note also that if *all* the data above the classifier was considered when making the updates, following similar arguments to those in the proof of Theorem 4.1, the algorithm would obtain overestimates of the distributions. Lastly, we could equivalently balance data by resampling the exploration data (rather than downsampling the exploitation data), to debias data through this procedure.

B.3. Proof of Theorem 4.3

Proof. We detail the proof for debiasing $\hat{f}_t^0$ (which happens using $x^\dagger \geq \text{LB}_t$ and $y^\dagger = 0$); the proof for $\hat{f}_t^1$ is similar.

Part (a). In time step $t+1$, with the arrival of a batch of N_{t+1} samples in $[\text{LB}_t, \infty)$, the current estimate $\hat{\omega}_t^0$ will be updated to $\hat{\omega}_{t+1}^0$ based on the proportion of $\hat{\omega}_t^0$ in the existing data. Denote the current left portion in $(\text{LB}_t, \hat{\omega}_t^0)$ as $p_1 := \frac{\hat{F}^0(\hat{\omega}_t^0) - \hat{F}^0(\text{LB}_t)}{\hat{F}^0(\theta_t) - \hat{F}^0(\text{LB}_t)}$. Based on Definition 3.1, we can also obtain the portion in $(\hat{\omega}_t^0, \theta_t)$ denoted as $p_2 := \frac{\hat{F}^0(\theta_t) - \hat{F}^0(\hat{\omega}_t^0)}{\hat{F}^0(\theta_t) - \hat{F}^0(\text{LB}_t)} = p_1$. Since their denominators are the same, we can just compare their numerators. Denote μ^0 and $\hat{\mu}_t^0$ as the mode of the true and estimated label 0 distribution, where μ^0 is unknown and $\hat{\mu}_t^0$ is the current estimates at time step t. When $\hat{\mu}_t^0 < (resp. >)\mu^0$, it is underestimated (resp. overestimated), and w.l.o.g we can assume the $\hat{\omega}_t^0 < (resp. >)\omega^0$. For example, in Gaussian and Beta distribution with α parameter unknown. Denote $K = 2F^0(\hat{\omega}_t^0) - F^0(\text{LB}_t) - F^0(\theta_t)$. Notice that the only unknown variable in K is the unknown distribution parameter hidden in F^0, which can be written as a function of mode μ^0. Hence, we can write K as a function of μ^0. For example, in Gaussian distribution, μ^0 is the mean of the distribution; in Beta distribution, the unknown parameter α can be written in terms of the mode μ^0 given parameter β is known. Based on Definition 3.1, $K = 0$ when $\mu^0 = \hat{\mu}_t^0$, which represents the perfectly estimated case ($\hat{\omega}_t^0 = \omega^0$). In this case, $p_1 = p_2$, which means once the parameter is correctly estimated, $\hat{f}_t^0$ is not expected to shift from f^0.

Now, we can find the sign of K when μ^0 is above or below the $\hat{\mu}_t^0$ by taking the derivative of K w.r.t. μ^0. Denote K' as

$$\begin{aligned} K' = \frac{dK}{d\mu^0} &= \frac{d}{d\mu^0}\left[2F^0(\hat{\omega}_t^0) - F^0(\text{LB}_t) - F^0(\theta_t)\right] \\ &= -2f_{\mu^0}^0(\hat{\omega}_t^0) + f_{\mu^0}^0(\text{LB}_t) + f_{\mu^0}^0(\theta_t) \end{aligned}$$

The expression $f_{\mu^0}^0(\hat{\omega}_t^0)$ means the distribution pdf calculated at $\hat{\omega}_t^0$ with unknown variable μ^0. The last equality holds true because $F^0(\hat{\omega}_t^0)$ becomes smaller when μ^0 becomes larger while $\hat{\omega}_t^0$ holding as constant. For example, in Gaussian distribution $F^0(\hat{\omega}_t^0)$ in $N(6,1)$ is larger than that in $N(7,1)$ for any given $\hat{\omega}_t^0$.

Notice that, $\hat{\omega}_t^0$ is the α-th percentile of the $\hat{f}_t^0$, which is a random but known constant. When $\mu^0 = \hat{\omega}_t^0$, since we have $\text{LB}_t \leq \hat{\omega}_t \leq \theta_t$ based on Definition 3.1, we can find $K'(\hat{\omega}_t^0) \leq 0$ since the density at mode μ^0 is the largest. Hence, we can relate the location of the $\hat{\omega}_t^0$ with μ^0 together and conclude $K' \leq 0$ for any μ^0. Therefore, we consider the following two cases:

Case 1 (Underestimated): $\hat{\omega}_t^0 < \omega^0$. Based on our assumption, we can also have $\hat{\mu}_t^0 \leq \mu^0$. Since $K = 0$ when $\mu^0 = \hat{\mu}_t^0, \hat{\omega}_t^0 = \omega^0$, and $K' \leq 0$ for any μ^0. Then, in this case, we have $K \leq 0$ when $\hat{\mu}_t^0 \leq \mu^0$, which means $p_1 \leq p_2$. Therefore, more samples are expected to be observed in range of $(\hat{\omega}_t^0, \theta_t)$ so that the $\hat{\omega}_t^0$ is expected to shift up. Hence, we have $\mathbb{E}[\hat{\omega}_{t+1}^0] \geq \hat{\omega}_t^0$.

Case 2 (Overestimated): $\omega^0 < \hat{\omega}_t^0$. Through similar analysis as Case 1 (Underestimated), we can obtain $\mathbb{E}[\hat{\omega}_{t+1}^0] \leq \hat{\omega}_t^0$.

Part (b). We first show that the converging sequence converges to the true estimates.

By the construction of the bounds in Definition 3.1, the estimated parameter $\hat{\omega}_t^0$ is the α-th percentile of $\hat{f}_t^0$, the median in the interval $[\text{LB}_t, \theta_t]$ and some percentile in the interval $[\text{LB}_t, \infty)$; we therefore first find their distribution accordingly. Assume there are $N_t = m + n + 1$ points in the interval $[\text{LB}_t, \infty)$ with m and n samples below and above $\hat{\omega}_t^0$ respectively. More specifically, for these n samples, there are m samples between $[\hat{\omega}_t^0, \theta_t]$ and $n - m$ samples above θ_t. Based on the probability distribution of order statistics in $[\text{LB}_t, \theta_t]$, denote three possibilities X, Y, Z denoting the number of samples below, on, and above the $\hat{\omega}_t^0$, respectively, having probabilities $p = \frac{F^0(\hat{\omega}_t^0) - F^0(\text{LB}_t)}{F^0(\theta_t) - F^0(\text{LB}_t)}$, $q = \frac{f^0(\hat{\omega}_t^0)}{F^0(\theta_t) - F^0(\text{LB}_t)}$, and $r = \frac{F^0(\theta_t) - F^0(\hat{\omega}_t^0)}{F^0(\theta_t) - F^0(\text{LB}_t)}$. Since

the distributions are continuous, the probability of multiple samples being exactly on $\hat{\omega}_t^0$ is zero. Therefore, the pdf of $\hat{\omega}_t^0$ can be found based on the density function of the trinomial distribution:

$$\mathbb{P}(\hat{\omega}_t^0 = \nu)\mathrm{d}\nu = \frac{(2m+1)!}{m!m!}\left(\frac{F^0(\nu)-F^0(\mathrm{LB}_t)}{F^0(\theta_t)-F^0(\mathrm{LB}_t)}\right)^m\left(\frac{F^0(\theta_t)-F^0(\nu)}{F^0(\theta_t)-F^0(\mathrm{LB}_t)}\right)^m\frac{f^0(\nu)}{F^0(\theta_t)-F^0(\mathrm{LB}_t)}\mathrm{d}\nu \tag{2}$$

From the above, we can see that the density function of the $\hat{\omega}_t^0$ is a beta distribution with $\alpha = m+1, \beta = m+1$, pushed forward by $H(\nu) := \frac{F^0(\nu)-F^0(\mathrm{LB}_t)}{F^0(\theta_t)-F^0(\mathrm{LB}_t)}$; this is the CDF of the truncated F^0 distribution in $[\mathrm{LB}_t, \theta_t]$. In other words, using G to denote the Beta distribution's CDF, $\hat{\omega}_t^0$ has CDF $G(H(\nu))$, and by the chain rule, pdf $g(H(\nu))h(\nu)$.

It is known (Maritz & Jarrett, 1978) that for samples located in the range of $[\mathrm{LB}_t, \theta_t]$, the sampling distribution of the median becomes asymptotically normal with mean M and variance $\frac{1}{4(2m+3)H(M)}$, where M is the median, the truncated F^0 distribution in $[\mathrm{LB}_t, \theta_t]$. If the sequence of $\{\hat{\omega}_t^0\}$ produced by our `active debiasing` algorithm converges, by Definition 3.1, the thresholds LB_t and θ_t will converge as well; As $t \to \infty$, $\epsilon_t \to 0$, $2m+1 \to \infty$ in this interval, the variance becomes zero, and $\hat{\omega}_{t+1}^0 \to M$. By Definition 3.1, it must be that the median M of H is equal to ω^0. Therefore, $\hat{\omega}_{t+1}^0 \to \omega^0$.

Lastly, we show that the sequence of estimates $\{\hat{\omega}_t^0\}$ is a converging sequence. Consider the sequence of estimates $\{\hat{\omega}_t^0\}$, and separate into the two disjoint subsequences $\{\hat{y}_t^0\}$ denoting the parameters that are underestimated with respect to the true ω^0, and $\{\hat{z}_t^0\}$ denoting those that are overestimated.

We now show that the sequence of underestimation errors, $\{\Delta_t^y\} := \{\omega^0 - \hat{y}_t^0\}$ and the sequence of overestimation errors, $\{\Delta_t^z\} := \{\hat{z}_t^0 - \omega^0\}$, are supermartingales. We detail this for $\{\Delta_t^y\}$. Consider two cases:

- First, assume the update $\hat{y}_{t+1}^0$ is the next immediate update after $\hat{y}_t^0$ in the original sequence $\{\hat{\omega}_t^0\}$; that is, an underestimated $\hat{y}_t^0$ has been updated to a parameter that continues to be an underestimate. In this case, by Part (a), $\mathbb{E}[\hat{y}_{t+1}^0|\hat{y}_t^0] \geq \hat{y}_t^0$, and therefore, $\mathbb{E}[\Delta_{t+1}^y|\Delta_t^y] \leq \Delta_t^y$.
- Alternatively assume $\hat{y}_{t+1}^0$ is not obtained immediately from $\hat{y}_t^0$; that is, $\hat{y}_{t+1}^0$ has been obtained as a result of an update from an overestimated parameter. We note that now, $\hat{y}_{t+1}^0 \geq \hat{y}_t^0$. This is because either no new estimates have been obtained between $\hat{y}_t^0$ and the true parameter ω^0 since the last time the parameter was underestimated, in which case, it must be that $\hat{y}_{t+1}^0 = \hat{y}_t^0$. Otherwise, a new estimate in $[\hat{y}_t^0, \omega^0]$ has been obtained, in which case, again, $\mathbb{E}[\hat{y}_{t+1}^0|\hat{y}_t^0] \geq \hat{y}_t^0$. In either case, $\mathbb{E}[\Delta_{t+1}^y|\Delta_t^y] \leq \Delta_t^y$.

Therefore, by the Doobs Convergence theorem, the supermatingales $\{\Delta_t^y\}$ and $\{\Delta_t^z\}$ converge to random variables Δ^y and Δ^z. By the same argument as the beginning of the proof of this part, these are asymptotically normal with mean zero and with variances decreasing in the number of observed samples in their respective intervals. Therefore, $\Delta^y \to 0$ and $\Delta^z \to 0$ as $N \to \infty$, and therefore $\{\hat{\omega}_t^0\}$ converges to ω. □

B.4. Proof of Theorem 4.4

Proof. This proof is based on a reduction from fair-classification to a sequence of cost-sensitive classification problems, as proposed and also used to obtain error bounds in Agarwal et al. (2018) and Bechavod et al. (2019). We adapt these to our bounded exploration setting. In order to find out our algorithm's error bound, we will have four steps. The first step is to rewrite each individual update as saddle point problem, which can be solved efficiently by exponentiated gradient reduction method introduced in Agarwal et al. (2018). Second, based on the solution output from the reduction method, we find the bound of classification error on the true distribution. Thirdly, we will introduce the exploration error made by our debiasing algorithm. Lastly, we will aggregate m updates together to derive the final algorithm error bound. For a simpler notation, only one group will be studied and subscript $t = 1$ will be assigned in the following step 1-3 since they focus on the information in the first update, and the other group can be analyzed in the same way. We also assume throughout that a fairness constraint $|\mathcal{C}(\theta_{a,t}, \theta_{b,t})| \leq \gamma$ has been imposed.

Step 1: Rewrite. We will treat our first update as a saddle point problem. Denote $err(h_{\theta_{g,t=1}}) = \sum_{i=1}^{b_{0a}+b_{1a}+b_{0b}+b_{1b}|1} \mathbb{E}_{(x_i,y_i,g_i)\sim D}\left[\ell(h_{\theta_{g,t=1}}(x_i, g_i), y_i)\right]$. Since we do not know the true distribution over (X, Y, G), and only have access to samples, we will use the empirical estimates $\hat{err}(h_{\theta_{g,1}})$ and $\hat{\mathcal{C}}(\theta_{a,1}, \theta_{b,1})$. Due to the sampling error, we

also allow errors in satisfying the constraints by setting $\hat{\gamma} = \gamma + e$. For the fairness constraint, we will introduce Lagrangian multipliers $\lambda_j \geq 0$. This allows us to define the Lagrangian of the problem:

$$\mathcal{L}(h_{\theta_{g,1}}, \lambda_j) = e\hat{r}r(h_{\theta_{g,1}}) + \lambda_1(\hat{\mathcal{C}}(\theta_{a,1}, \theta_{b,1}) - \hat{\gamma}) + \lambda_2(-\hat{\mathcal{C}}(\theta_{a,1}, \theta_{b,1}) - \hat{\gamma})$$

Following the proof procedures in Agarwal et al. (2018), we impose an additional constraint on the l_1 norm of λ_j such that $||\boldsymbol{\lambda}||_1 \leq B$ for a sufficient large constant B. By strong duality, we have:

$$OPT = \min_{\theta_{g,1} \in \mathbb{R}_+} \max_{\lambda_j \in \mathbb{R}_+, ||\boldsymbol{\lambda}||_1 \leq B} \mathcal{L}(h_{\theta_{g,1}}, \lambda_j) = \max_{\lambda_j \in \mathbb{R}_+, ||\boldsymbol{\lambda}||_1 \leq B} \min_{\theta_{g,1} \in \mathbb{R}_+} \mathcal{L}(h_{\theta_{g,1}}, \lambda_j)$$

where the optimal solution $(h^*_{\theta_{g,1}}, \lambda^*_j)$ is the saddle point of the $\mathcal{L}(h_{\theta_{g,1}}, \lambda_j)$, and OPT denotes the optimal objective value. Agarwal et al. (2018) proposed an exponentiated gradient algorithm to find an approximate solution corresponding to a v-approximate saddle point $(\hat{h}_{\theta_{g,1}}, \hat{\lambda}_j)$ of the Lagrangian such that:

$$\mathcal{L}(\hat{h}_{\theta_{g,1}}, \hat{\lambda}_j) \leq \mathcal{L}(h_{\theta_{g,1}}, \hat{\lambda}_j) + v \text{ for all } \theta_{g,1} \in \mathbb{R}_+$$

$$\mathcal{L}(\hat{h}_{\theta_{g,1}}, \hat{\lambda}_j) \geq \mathcal{L}(\hat{h}_{\theta_{g,1}}, \lambda_j) - v \text{ for all } \lambda_j \in \mathbb{R}_+, ||\boldsymbol{\lambda}||_1 \leq B$$

Hence, as shown in their Theorem 1, we can also find a v-approximate saddle point in at most $O(1/v^2)$ iterations. However, a large value of B will increase the cost of needing more iterations to reach any given suboptimality. Hence, following from Lemma B.3 of Bechavod et al. (2019), they show that it is sufficient to reduce it to be $\Lambda = \{||\boldsymbol{\lambda}||_1 \leq 2|\boldsymbol{\lambda} \in \mathbb{R}^2_+\}$. We also adopt this assumption.

Step 2: Bound of error on the true distribution.

Lemma B.1 (Follows from Lemma B.4 of Bechavod et al. (2019))**.** *Suppose $(\hat{h}_{\theta_{g,1}}, \hat{\lambda}_j)$ is a v-approximate saddle point of the Lagrangian. Then, the following inequality holds:*

$$e\hat{r}r(\hat{h}_{\theta_{g,1}}) \leq e\hat{r}r(h^*_{\theta_{g,1}}) + 2v$$

The algorithm error comes with three different sources. First, we use samples to estimate the true distribution. Second, we introduce a bound B on the magnitude of $\boldsymbol{\lambda}$. Lastly, we have the suboptimal solution that is returned by the exponentiated gradient algorithm with suboptimiality level v. The first error is unavoidable, which is also called the statistical error. The other two can be driven arbitrarily small at the cost of more iterations of exponentiated gradient algorithm. To bound the statistical error, we use Rademacher complexity of the classifier family $\mathcal{H}$ denoted as $\mathcal{R}_n(\mathcal{H})$, where n is the number of training samples. Denote $n'_{g,t}$ be the number of training samples we collected in round t in group g. In the first update, we have $n'_{g,1} = b_{0g} + b_{1g}|\{t = 1\}$. We also assume that $\mathcal{R}_n(\mathcal{H}) \leq Cn^{-\alpha}$ for some $C \geq 0$ and $\alpha \leq 1/2$. Hence, based on the Theorem 4 in Agarwal et al. (2018), we can find that in the first update with probability at least $1 - 4\delta$ with $\delta > 0$:

$$err(\hat{h}_{\theta_{g,1}}) \leq err(h^*_{\theta_{g,1}}) + 2v + 4\mathcal{R}_{n'_{g,1}}(\mathcal{H}) + \frac{4}{\sqrt{n'_{g,1}}} + \sqrt{\frac{2\ln(2/\delta)}{n'_{g,1}}}$$

Step 3: Introduce exploration error. Let $n_{0,g,t}$ and $n_{1,g,t}$ denote the number of samples from un/qualified group that fall below the threshold θ_t in round t respectively. Since in Step 2 we already considered the classification errors, we only consider the additional exploration error introduced in order to remove biases. Because of exploration, some samples from the qualified group rejected previously will be accepted, which will allow the algorithm to make less errors. And the same situation also happens to the unqualified group, which will make more errors.

Denote ϵ_t as the exploration probability at round t. Hence, for the pure exploration model, we should add errors made for unqualified group and minus correct decisions made for qualified group. Mathematically, it can be represented as:

$$(n_{0,g,t} - n_{1,g,t})\epsilon_t \mathbb{1}\Big[(x_i|g_i) \leq \theta_{g,t}\Big]$$

Comparing to our bounded exploration model, we introduce the LB_t to limit the depth of exploration. In other words, the number of samples fall into the exploration range will be proportional to $n_{0,g,t}$ and $n_{1,g,t}$ based on the location of LB_t. Mathematically, denote $N_{g,t}$ as the net exploration error for group g at round t such that:

$$N_{g,t} = \left(\frac{\hat{F}^0(\theta_t) - \hat{F}^0(\text{LB}_t)}{\hat{F}^0(\theta_t)}\epsilon_t n_{0,g,t} - \frac{\hat{F}^1(\theta_t) - \hat{F}^1(\text{LB}_t)}{\hat{F}^1(\theta_t)}\epsilon_t n_{1,g,t}\right)\mathbb{1}\Big[LB_{g,t} \leq (x_i|g_i) \leq \theta_{g,t}\Big]$$

Note that, when the $\text{LB}_t \to -\infty$, then the exploration error of our bounded exploration model is the same as pure exploration baseline.

Step 4: Add m updates together. For the advantaged group a, by considering the exploration error and adding total m updates, we have

$$\sum_{t=1}^{m} err(\hat{h}_{\theta_{a,t}}) \leq \sum_{t=1}^{m} \left[err(h^*_{\theta_{a,t}}) + 4\mathcal{R}_{n'_{a,t}}(\mathcal{H}) + \frac{4}{\sqrt{n'_{a,t}}} + \sqrt{\frac{2\ln(2/\delta)}{n'_{a,t}}} + N_{a,t} \right] + 2mv$$

The same expression can also be obtained for the disadvantaged group b. Hence, let $n'_t := \min\{n'_{a,t}, n'_{b,t}\}, N_t := \max\{N_{a,t}, N_{b,t}\}, \mathcal{R}_{n''_t}(\mathcal{H}) := \max\{\mathcal{R}_{n'_{a,t}}(\mathcal{H}), \mathcal{R}_{n'_{b,t}}(\mathcal{H})\}$ and by adding these two expressions together, it yields the error bound for our algorithm such that

$$\begin{aligned}
\text{Error Bound} &= \sum_{t=1}^{m} err(\hat{h}_{\theta_{a,t}}) + \sum_{t=1}^{m} err(\hat{h}_{\theta_{b,t}}) - \sum_{t=1}^{m} err(h^*_{\theta_{a,t}}) - \sum_{t=1}^{m} err(h^*_{\theta_{b,t}}) \\
&\leq \sum_{g} \sum_{t=1}^{m} \left[4\mathcal{R}_{n'_{g,t}}(\mathcal{H}) + \frac{4}{\sqrt{n'_{g,t}}} + \sqrt{\frac{2\ln(2/\delta)}{n'_{g,t}}} + N_{g,t} \right] + 4mv \\
&\leq 8m\mathcal{R}_{n''_1}(\mathcal{H}) + \frac{8m}{\sqrt{n'_1}} + 2m\sqrt{\frac{2\ln(2/\delta)}{n'_1}} + 2mN_1 + 4mv
\end{aligned}$$

The last equality holds since groups are independent from each other, and the first update comes with the largest error. □

B.5. Proof of Theorem 4.5

Proof. We prove the proposition for the case where the introduction of fairness constraints leads to over-selection of group g, i.e., $\theta^F_{g,t} < \theta^U_{g,t}$. The proofs for the under-selected case are similar. We note that the presence of two different groups only affects the choice of the classifier given the fairness constraints, following which the proof becomes independent of the group label; we therefore drop g in the remainder of the proof.

We detail the proof for the debiasing of $\hat{f}^0_t$, which depends on the choice of LB_t in Definition 3.1, i.e.,

$$\hat{F}^0_t(\text{LB}_t) = 2\hat{F}^0_t(\hat{\omega}^0_t) - \hat{F}^0_t(\theta_t) \, .$$

Since $\theta^F_t < \theta^U_t$, this means that $\hat{F}^0_t(\theta^F_t) < \hat{F}^0_t(\theta^U_t)$, and consequently that $\hat{F}^0_t(\text{LB}^F_t) > \hat{F}^0_t(\text{LB}^U_t)$, and thus, that $\text{LB}^F_t > \text{LB}^U_t$.

Now, consider the interval $[\text{LB}_t, \max^0]$, with $\max^0$ denoting the maximum of f^0. For example, $\max^0$ can be ∞ if f^0 follows a Gaussian distribution; and it can be finite if f^0 follows a Beta distribution. Only arrivals of $(x^\dagger, y^\dagger)$, with $y^\dagger = 0$, who are admitted in this interval, will result in an update to the estimated median. Since $\text{LB}^F_t > \text{LB}^U_t$, this interval is narrower under the fairness constrained classifier, meaning that it takes more time to meet the batch size requirement under compared LB^U_t compared to LB^F_t. As detailed in the proof of Theorem 4.3 each of these updates will move the estimate in the correct direction, and these estimates converge to the true value in the long-run as more samples become available. Hence, debiasing of $\hat{f}^0_t$ is slower after the introduction of fairness constraints.

Similar arguments hold for updating $\hat{f}^1_t$, which takes samples in $[\text{LB}_t, \max^1]$. When LB_t increases, it also takes more time for label 1 distribution update. Hence, after the introduction of the constraint, the fairness unconstrained classifier observes a wider range of samples points, including all those observed by the constrained classifier. Therefore, the addition of fairness constraints decreases the speed of debiasing on $\hat{f}^1_t$ as well. □

C. Additional Figures and Experiments

C.1. Performance of `active debiasing` on Beta distributions

Fig. 5 shows that our algorithm can debias data for which the underlying feature-label distributions follow Beta distributions. We have assumed a mistmach between the parameter α of the true and estimated distributions, and selected these so that the estimated and true distributions have different relative skewness. This verifies that Theorem 4.3 can hold beyond symmetric distributions.

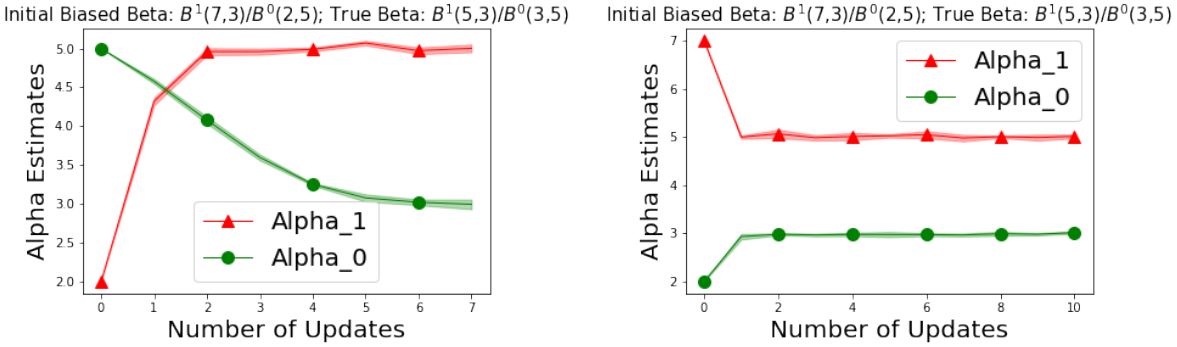
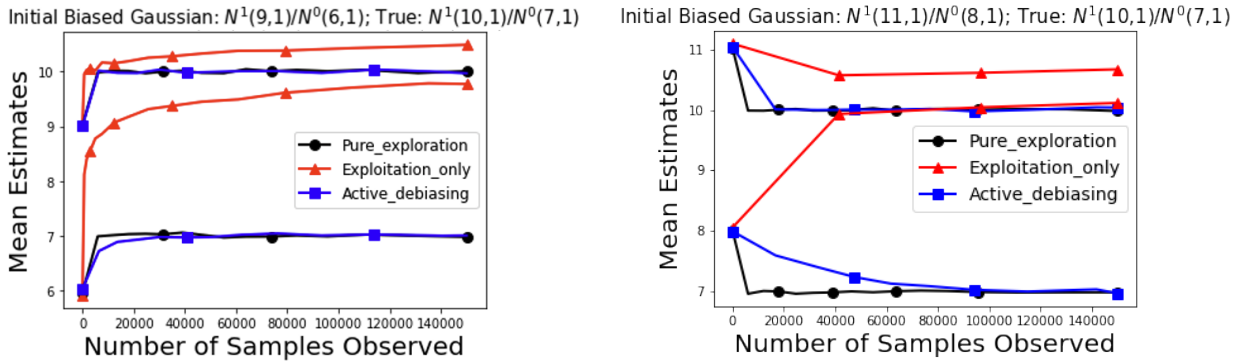


Figure 5. `Debiasing` under Beta distributions.

C.2. Comparison with the `exploitation-only` and `pure exploration` baselines

Fig. 6 compares our algorithm against two baselines. The underlying distributions are Gaussian and no fairness constraint is imposed. Our algorithm sets $\alpha^1 = 50$ and $\alpha^0 = 60$ percentiles, and exploration frequencies ϵ_t are selected adaptively by both our algorithm and pure exploration.



(a) Rate of debiasing, f^1 and f^0 underestimated.

(b) Rate of debiasing, f^1 and f^0 overestimated.

Figure 6. Speed of debiasing of `active debiasing` vs. `exploitation-only` and `pure exploration`

C.3. `Active debiasing` on the *Adult* dataset

Fig. 7 illustrates the performance of our algorithm on the *Adult* dataset. Data is grouped based on race (White G_a and non-White G_b), with labels $y = 1$ for income $> \$50k/year$. A one-dimensional feature $x \in \mathbb{R}$ is constructed by conducting logistic regression on four quantitative and qualitative features (education number, sex, age, workclass), based on the initial training data.[1] Using an input analyzer, we found Beta distributions as the best fit to the underlying distributions. We use 2.5% of the data to obtain a biased estimate of the parameter α. The remaining data arrives sequentially. We use $\alpha^1 = 50$ and $\alpha^0 = 60$ and a fixed decreasing $\{\epsilon_t\}$, with the equality of opportunity fairness constraint imposed throughout.

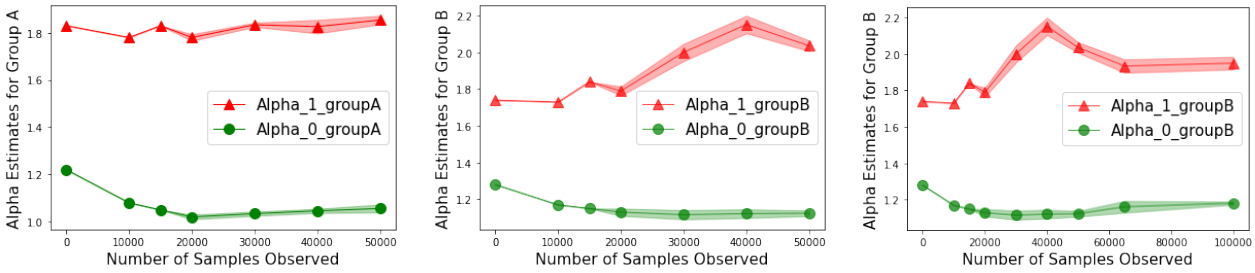
C.4. `Active debiasing` on the *FICO* dataset

Fig. 8 also illustrates the performance of our algorithm on the *FICO* dataset (Reserve, 2007; Hardt et al., 2016), and shows that it is successful in both groups and on both labels.

C.5. Additional experiments on the impact of depth of exploration

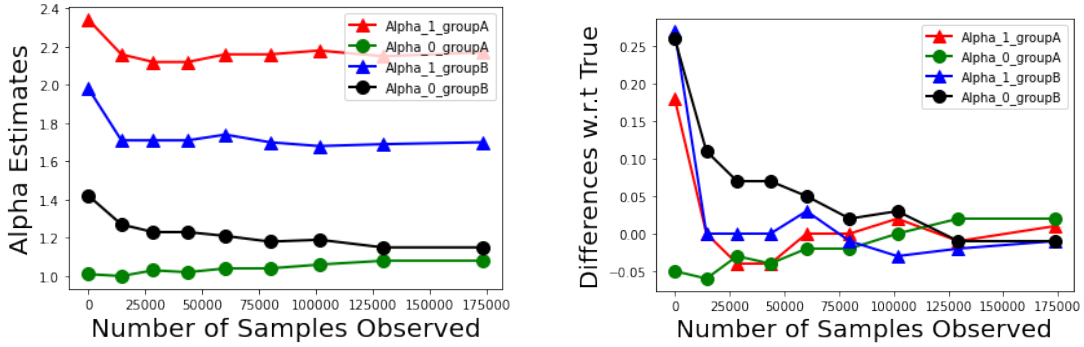
Figure 9 compares the effects of modifying the depth of exploration through the choice of reference points on the performance of our `active debiasing` algorithm. In particular, we fix $\alpha^1 = 50$ as the reference point on the qualified agents'

[1] While this experiment maintains the same mapping throughout, the mapping could be periodically revised.



(a) Debiasing G_a. (b) Debiasing G_b. (c) *Adult* dataset augmented with 60,000 synthetic samples.

Figure 7. Illustration of the performance of `active debiasing` on the *Adult* dataset. The true underlying distributions were estimated to be Beta distributions with parameters Beta(1.94, 3.32) and Beta(1.13, 4.99) for group a (White) label 1 and 0, respectively, and Beta(1.97, 3.53) and Beta(1.19, 6.10) for group b (non-White) label 1 and 0, respectively. We used 2.5% of the data to fit initial assumed distributions Beta(1.83, 3.32) and Beta(1.22, 4.99) for group a label 1 and 0, respectively, and Beta(1.74, 3.53) and Beta(1.28, 6.10) for group b label 1 and 0, respectively. The equal opportunity fairness constraint is imposed throughout. The exploration frequency $\{\epsilon_t\}$ is reduced with the fixed schedule of being subtracted by 0.1 after observing every 10000 samples



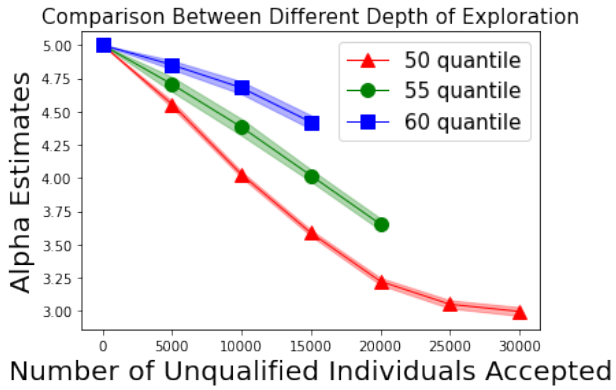
(a) Active Debiasing on the *FICO* dataset. (b) Difference w.r.t. the true value.

Figure 8. Illustration of the performance of `active debiasing` on the *FICO* dataset. The true underlying distributions were estimated to be Beta distributions with parameters Beta(2.16, 1.27) and Beta(1.06, 3.98) for group a (White) label 1 and 0, respectively, and Beta(1.71, 1.62) and Beta(1.16, 5.51) for group b (non-White) label 1 and 0, respectively. We used 0.3% of the data to fit initial assumed distributions Beta(2.34, 1.27) and Beta(1.01, 3.98) for group a label 1 and 0, respectively, and Beta(1.98, 1.62) and Beta(1.42, 5.51) for group b label 1 and 0, respectively. The equal opportunity fairness constraint is imposed throughout. The exploration frequency $\{\epsilon_t\}$ is reduced with the fixed schedule of being subtracted by 0.1 after observing every 17000 samples

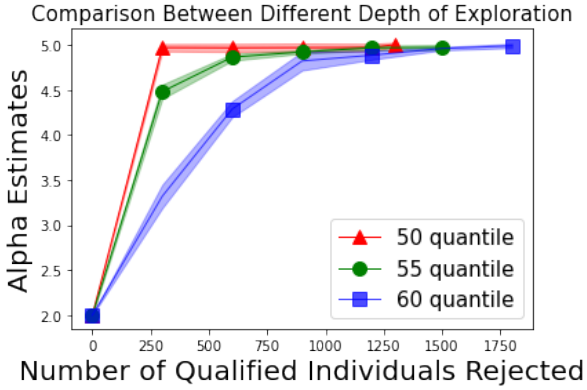
estimates, and vary the reference points on unqualified agents' estimates in $\alpha^0 \in \{50, 55, 60\}$, with smaller reference points indicating deeper exploration (see Definition 3.1). In all three settings, we reduce $\{\epsilon_t\}$ following a fixed reduction schedule, as described in Section 5.

We first note that as also observed earlier, increasing the depth of exploration (here, e.g., setting $\alpha^0 = 50$) leads to faster speed of debiasing. This additional speed comes with a tradeoff: Fig. 9(a) shows that algorithms with deeper exploration make more false positive errors, as they accept more unqualified individuals during exploration; by taking on this additional risk, they can debias the data faster. In addition, as observed in Fig. 9(b), the increased speed of debiasing means that the algorithm ultimately ends up making *fewer* false negative decisions on the qualified individuals as a result of obtaining better estimates of their distributions.

We conclude that a decision maker can use the choice of the reference point α^0 in our proposed algorithm to achieve their preferred tradeoff between the risk incurred due to incorrect admissions (higher FP) vs the benefit from the increased speed of debiasing and fewer missed opportunities (fewer FN).



(a) False positives (unqualified agents admitted) under each reference point



(b) False negatives (qualified agents rejected) under each reference point

Figure 9. `Active debiasing` under different choices of depth of exploration, with $\alpha^1 = 50$ and $\alpha^0 = \{50, 55, 60\}$. We reduce $\{\epsilon_t\}$ following a fixed reduction schedule. The underlying feature distributions are Beta distributions.

C.6. Debiasing with two unknown parameters: a Gaussian distribution with two unknown parameters mean μ and variance σ^2

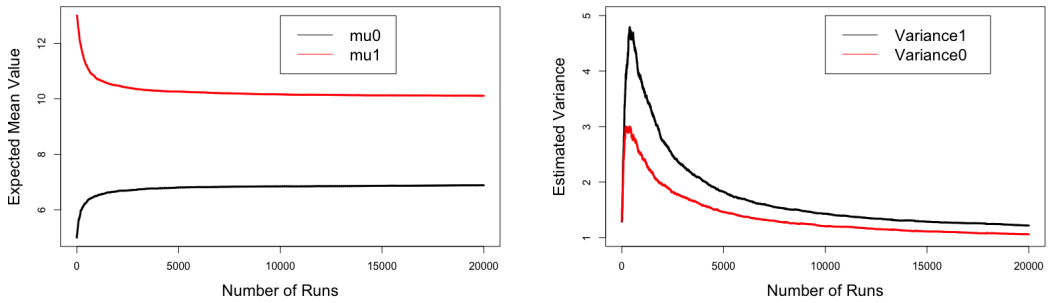
In this subsection, we extend our algorithm to debias the estimates of distributions with two unknown parameters. Specifically, we consider a single group, and assume that the underlying feature-label distributions are Gaussian distributions for which both the mean and variance are potentially incorrectly estimated by the firm.

To help our analysis and simplify the experiment setting, similar to the LB_t setting in Definition 3.1, we can find a corresponding UB_t such that

$$\text{UB}_t = (\hat{F}_t^1)^{-1}(2\hat{F}_t^1(\hat{\omega}_t^1) - \hat{F}_t^1(\text{LB}_t))$$

where LB_t is obtained from Definition 3.1, $\hat{F}_t^0$, $(\hat{F}_t^1)^{-1}$ are the cdf and inverse cdf of the estimates $\hat{f}_t^1$, respectively, and $\hat{\omega}_t^1$ is (wlog) the α-th percentile of $\hat{f}_t^1$. We follow our `active debiasing` algorithm, with a choice of medians as reference points (i.e., $\alpha^i = 50, \forall i$), and setting the thresholds LB_t and UB_t so that the reference points are the medians of the truncated distribution between the bounds and the classifier θ_t. We then follow Algorithm 1's procedure with the same type of exploitation and exploration decisions, and with the additional step that now we update both parameters when updating the underlying estimates.

In order to update the mean and variance estimates for obtaining $\hat{f}_t^i$, we find the sample mean and sample variance of the collected data, incrementally. However, we note that the obtained sample mean and sample variances are *for truncated distributions*; the truncations are due to the presence of a classifier which limits the admission of a samples, as well as due to our proposed bounds LB_t and UB_t in the data collection procedure. We therefore need to convert between the estimated



(a) Debiasing the means. (b) Debiasing the variances.

Figure 10. Debiasing algorithm when both mean and variance of a Gaussian distribution are incorrectly estimated. The true underlying distributions are $f^1 \sim N(10, 1)$ and $f^0 \sim N(7, 1)$, and the initial estimates are $\hat{f}_0^1 \sim N(13, 1.3)$ and $\hat{f}_0^0 \sim N(5, 1.3)$. The algorithm corrects both biases in the long run.

statistics for the truncated distribution and those of the full distribution accordingly.

Specifically, we obtain the sample mean of the truncated distribution as follows:

$$\hat{\mu}_{t+1}^i = \frac{x_1 + x_2 + ... + x_{n_i} + x^\dagger}{N_t^i + 1} = \frac{N_t^i}{N_t^i + 1}\hat{\mu}_t^i + \frac{x^\dagger}{N_t^i + 1}, \quad i \in \{0, 1\} .$$

where N_t^i is the existing number of agents in the pool, and μ_t^i is the current (truncated) mean value estimate for label $i = \{0, 1\}$.

For the sample (truncated) variance for group i, $(\hat{s}_t^i)^2$, the updating procedure is:

$$\begin{aligned}(\hat{s}_{t+1}^i)^2 &= \frac{\sum_{j=1}^{N_t^i}(\hat{\mu}_t^i - x_j)^2 + (\hat{\mu}_t^i - x^\dagger)^2}{N_t^i + 1 - 1} \\ &= \frac{\sum_{j=1}^{N_t^i} x_j^2 + (x^\dagger)^2 - (N_t^i + 1)(\hat{\mu}_t^i)^2}{N_t^i + 1 - 1} \\ &= \frac{N_t^i - 1}{N_t^i}(\hat{s}_t^i)^2 + \frac{(x^\dagger)^2 - (\hat{\mu}_t^i)^2}{N_t^i}, \quad i \in \{0, 1\} .\end{aligned}$$

After finding the above estimates of the mean and variance of the truncated distribution, we need to estimate the mean and variance of the *full* underlying distribution. We first note that given our choice of bounds LB_t and UB_t, the mean of the underlying distribution is (assumed to be) the same as that of the truncated distribution. To find the untruncated variance for the full distribution, we use the following relation between the variances of truncated and untruncated Gaussian distributions:

$$Var(x|a \le x \le b) = s^2 = \sigma^2 \left[1 + \frac{\alpha\phi(\alpha) - \beta\phi(\beta)}{\Phi(\beta) - \Phi(\alpha)} - (\frac{\phi(\alpha) - \phi(\beta)}{\Phi(\beta) - \Phi(\alpha)})^2\right]$$

where $\alpha = \frac{a-\mu}{\sigma}$, $\beta = \frac{b-\mu}{\sigma}$, $\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ and $\Phi(x) = \frac{1}{2}(1 + erf(\frac{x}{\sqrt{2}}))$. In our algorithm, $a = \theta_t$ and $b = \text{UB}_t$ for $i = 1$, and $a = \text{LB}_t$ and $b = \theta_t$ for $i = 0$. We note that in both cases, we can drop the third term in the above formula since based on our algorithm, a, b are symmetric around the mean value, so that $\phi(\alpha) = \phi(\beta)$. We solve the above equations to find $\hat{\sigma}_t^i$ from the truncated estimates $\hat{s}_t^i$.

Figure 10 shows that the debiasing algorithm with the update procedures described above can debias both parameters in the long run. We do observe that the debiasing of the variance initially increases its error. This is because, initially, when observing samples outside of its believed range (due to a combination of incorrectly estimated means and variances), the algorithm increases its estimates of the variance to explain such samples. However, as the estimate of the mean is corrected, the variance can be reduced as well and become consistent with the collected observations. Ultimately, both parameters will be correctly estimated.