Appendix

A More on Limitations, Extensions, and Social Impacts

Limitations of Assumption 1, and its extensions. Our work, similar to the literature on multi-armed bandit learning, focuses on learning of a single unknown parameter (Assumption 1). This assumption has also been adopted in prior work on studying biases in adaptively collected data [31]. In addition, our problem is also similar to the commonly studied problem of estimating a *location parameter* of a probability distribution in statistics. Despite the commonality, this assumption entails parametric knowledge of the underlying distribution with the other parameters such as variance or spread being known (or if unknown, immaterial) to the algorithm. This may be a limit in the applicability of our algorithm.

As a first extension, in Appendix I, we have extended our algorithm and shown that it can debias the data in Gaussian distributions when both the mean and the variance of the distribution are unknown. Our algorithm still uses the same adaptive and bounded exploration mechanism outlined in Algorithm 1, and the LB set by Definition 1. In particular, the bounds are still chosen so as to balance the estimated reference point (i.e. location parameter) of the distribution around its true parameter once correctly estimated; nonetheless, we show the data collected in this way can also be used to debias the variance.

One potential drawback of this extension is that the debiasing of the variance is not *robust*, in that although the algorithm debiases both parameters in the long-run, it initially increases the error in the variance estimate. 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. The extension of our bounded exploration and update procedure to account for both mean and variance debiasing and use of more robust estimators remain as an interesting direction of future work.

Alternatively, the algorithm could use an input analyzer or distribution fit estimators of a given software to fit a desired distribution to the dataset available at the end of each round, by fixing the known parameters and fitting for the unknown parameters. Exploring this option, especially when additional parameters from the distribution are unknown, remains a direction of future research.

One-dimensional features and threshold classifiers. Our current analytical results have focused on one-dimensional feature data and threshold classifiers. Our experiments have also considered the use of our `active debiasing` algorithm on data with multi-dimensional features by performing a dimension reduction (e.g., as shown in the Adult dataset experiment).

In general, such dimension reductions may lead to loss of information. To evaluate this in our setting, we run experiments on the Adult dataset with classifiers trained with and without performing dimension reduction. The experiments show only minimal loss in accuracy, with results as follows:

• For the classifier trained through logistic regression without performing dimension reduction: the overall accuracy is 82.96%, and the accuracy for advantaged (41762 samples) and disadvantaged (7080 samples) groups are 82.13% and 88.26%, respectively.

• For the classifier trained through logistic regression after performing dimension reduction: The overall accuracy is 78.07%, and the accuracy for the advantaged and disadvantaged groups are 76.95% and 84.69%, respectively.

While this remains as a relative small ($\sim$ 5%) loss in performance, better dimension reduction approaches, or debiasing algorithms that do not require feature dimension reduction, remain as interesting directions of future work.

Extensions of our analytical results. We conjecture that Theorem 3 on the performance of `active debiasing` can be extended to distributions beyond unimodal distributions. Further, the analytical study of weighted regret of our algorithm, and comparison against the regret incurred by our two baselines, which we have observed numerically in Section 5, remain as main directions of future work. As another future extension, we also propose considering a mixture of normal distributions with varying means and fixed covariances as the assumed form of the underlying distributions that are

to be debiased. Such mixtures can be used to approximate other smooth probability density functions. An extension of our analytical results from unimodal distributions to these mixtures, together with extensions of our experimental results, could also address our algorithm's single unknown parameter limitation.

Potential (negative) social impacts. We have elaborated on the interplay of our algorithm with fairness constraints in Section 4.5 as well as in our experiments. Accounting for the different impact of fairness constraints on the speed of debiasing of different groups may be a consideration when choosing to supplement debiasing efforts with fairness constraints, or vice versa.

Also, a limitation of our algorithm for groups with smaller representation is discussed at the end of Section 5. In particular, we observed that in the *Adult* dataset, as limited data is available on qualified, disadvantaged agents, the estimates on this population is not fully debiased despite the other estimates having been almost debiased. In other words, our debiasing algorithm is most effective on obtaining correct estimates on populations with sufficiently high representation. This may still be indirectly beneficial to the underrepresented populations, as by having better estimates on the represented population, the algorithm can better assess and impose fairness constraints.

B Additional and Detailed Related Work

Data debiasing with censored and costly feedback. Our paper is most closely related to the works of [14, 4, 20, 6, 17], who have investigated the impacts of data biases on (fair) algorithmic decision making. Ensign et al. [14]'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. [4] and Kilbertus et al. [20] 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 [4] 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 [20] 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 [10, 31, 30, 41] 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. [10] 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. [31] study the problem of estimating statistical parameters from adaptively collected data. Their proposed adaptive data collection method, which also similar to ours (Assumption 1) is used for single-parameter estimation, is one of online debiasing; our proposed data collection methods however differ: 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 [31], to which their proposed debiasing algorithms based on data splitting and modified maximum likelihood estimators are applied.

While [10, 31] propose ex-post methods for debiasing adaptively collected data, Neel and Roth [30] 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 [30], account for the costs of exploration in our data collection procedure. The recent work of Wei [41] 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 [30, 41], 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 and Stangl [6] and Jiang and Nachum [17], also considers the interplay between algorithmic fairness rules and data biases. Blum and Stangl [6] 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. [16, 13, 8]. 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 [26, 42, 27]. Our work contributes to this line of research, by analyzing the long-term effects of imposing fairness constraints on data collection and debiasing efforts.

Biases in adaptively collected data: A few other works have, similar to our work, explored the question of biases induced by a decision rule on data collection. Deshpande et al. [10] 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. [31] study the problem of estimating statistical parameters from adaptively collected data. Their proposed adaptive data collection method, which also similar to ours (Assumption 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 [31], to which their proposed debiasing algorithms based on data splitting and modified maximum likelihood estimators are applied.

Active learning: Our work is also related to the active learning literature. Balcan et al. [3] study the sample complexity of labeled data required for active learning, Kazerouni et al. [19] propose an algorithm involving exploration and exploitation-based adaptive sampling, verifying it using simulations, Abernethy et al. [1] propose an active sampling and re-weighting technique by sampling from the worst off group at each step with the goal of building a computationally efficient algorithm with strong convergence guarantees to improve the performance on the disadvantage (highest loss) group while satisfying the notion of min-max fairness, Noriega-Campero et al. [32] propose an adaptive fairness approach, which adaptively acquires additional information according to the needs of different groups or individuals given information budgets, to achieve fair classification. Similar to the approaches of these papers, we also compensate for adaptive sampling bias through exploration (by admitting individuals who would otherwise be rejected). Comparing to all these literature, we

start with a biased dataset, and we primarily focus on recovering the true distribution by bounded exploration, accounting for the cost of exploration, avoiding the adaptive sampling bias, and consider fairness issues as orthogonal to our data collection procedure (and as such, can apply our procedure to debiasing the estimates on a single group).

Selective labeling bias: Lakkaraju et al. [22] address the problem of evaluating the performance of predictive model under the selective labeling problem. They propose a contraction technique to compare the performance of the predictive model and human judge while they are forced to have the same acceptance rate. De-Arteaga et al. [9] study the problems arising due to selective labeling. Similar to us, they propose a data augmentation scheme by adding more samples that would be more likely rejected (we refer to this as exploration) to correct the sample selection bias. Their proposed data augmentation technique is similar to our bounded exploration, but it differs in its selection of samples in that it adds samples that would be more likely to be rejected.

Fair learning: Kallus and Zhou [18] show that residual unfairness remains even after the adjustment for fairness when policies are learned from a biased dataset. They propose a re-weighting technique (similarly, re-weighing ideas are explored in [6] and [17]) to solve the residual unfairness issue while accounting for the censoring/adaptive sampling bias.

Online mean estimation: Compared to this literature, the main technical challenges of our proposed bounded exploration in online mean estimation is that it involves evaluating the behavior of statistical estimates based on data collected from a truncated distribution with *time-varying* truncation. More specifically, our data collection interval is bounded and truncated (which has been considered in some prior work on distribution/mean estimation as well, e.g., Lai and Ying [21]) but our exploration interval $[LB_t, \infty)$ is itself adaptive (which we believe is the main new aspect) and is what has motivated our analysis in a finite sample regime in Theorem 3. Our focus on the interplay of fairness constraints with online estimation efforts (Proposition 1) is also new compared to this existing literature.

Performative prediction: Finally, the recent line of work on performative prediction proposed by Perdomo et al. [35] 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.

C Pseudo-code for Algorithm 1

The pseudo-code is shown in Algorithm 1.

D Proof of Theorem 1

Proof. We detail the proof for label 0 estimates $\hat{\omega}_t^0$, and discuss two cases while assuming (wlog) that the unknown ω^y being estimated is the distribution's mean. 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 $\mathbb{E}[x_{t'}] = \mathbb{E}[X|X \geq \theta_{t'}]$.

Algorithm 1: Adaptive Algorithm on Real Data

Input: fairness constraint $\mathcal{C}(\theta_{a,t=0}, \theta_{b,t=0})$, initial fit portion P
Result: Fair classifier $\theta_{g,t}$
$t \leftarrow 0, \epsilon_{g,t=0} \leftarrow 1$
$\hat{\omega}^y_{g,t=0} \leftarrow (\hat{F}^y_{g,t=0})^{-1}(\alpha'^y_g)$ for $y \in \{0,1\}$
$LB_{g,t=0} \leftarrow (\hat{F}^0_{g,t=0})^{-1}(2\hat{F}^0_{g,t=0}(\hat{\omega}^0_{g,t=0}) - \hat{F}^0_{g,t=0}(\theta_{g,t=0}))$
while $i \leq N$ **do**
  Data_trun^y_g = [ ]
  portion_left^y_g = $\hat{F}^y_{g,t}(LB_t \leq x \leq \alpha'^y_g)/\hat{F}^y_{g,t}(LB_t \leq x)$
  $k \leftarrow 0$
  while $k \leq S$ *and* $i \leq N$ **do**
    for $g \in G = \{a, b\}$ **do**
      if $\theta_{g,t} \leq x$ **then**
        Decision $\leftarrow 1$ (accept)
      else if $LB_{g,t} \leq x \leq \theta_{g,t}$ *and* $rand() \leq \epsilon_{g,t}$ **then**
        Decision $\leftarrow 1$ (accept)
      else if $x \leq LB_{g,t}$ **then**
        Decision $\leftarrow 0$ (reject)
      else
        Decision $\leftarrow 0$ (reject)
      end
      Append x into Data^y_g if accepted, and **append** x into Data_trun^y_g only if $x \in (LB_{g,t}, \theta_{g,t})$, or $(\theta_{g,t}, \infty)$ with probability $\epsilon_{g,t}$
    end
    $i \leftarrow i + 1, k = \min(\text{len(Data\_trun}^y_a), \text{len(Data\_trun}^y_b))$
  end
  $t \leftarrow t + 1$
  $\hat{\omega}^y_{g,t} \leftarrow quantile(\text{Data\_trun}^y_g, \text{portion\_left}^y_g)$; `/* Update reference value using all collected samples from the batch */`
  Map back from $\hat{\omega}^y_{g,t}$ to the single unknown parameter in the estimated distribution $\hat{f}^y_{g,t}$
  Retrain the classifier, output new threshold $\theta_{g,t}$ and $LB_{g,t}$; `/* Update classifier using all collected samples so far */`
  Update $\epsilon_{g,t}$; `/* Can be FSR or adaptive */`
end

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

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

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 $\hat{\omega}^0_T$ 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, $\hat{\omega}^0_t$ will stay overestimated. The proof for $\hat{\omega}^1_t$ is similar. □

E Proof of Theorem 3

Proof. We detail the proof for debiasing $\hat{f}^0_t$ (which happens using $x^\dagger \geq \text{LB}_t$ and $y^\dagger = 0$); the proof for $\hat{f}^1_t$ 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 1, we can also obtain the portion in $(\hat{\omega}_t^0, \theta_t)$ denoted as $p_2 := \frac{\hat{F}^0(\theta_t) - \hat{F}^0(\hat{\omega}_t^0)}{\hat{F}^0(\theta_t) - \hat{F}^0(\text{LB}_t)} = p_1$. We consider the following cases:

Case 1 (Perfectly estimated): $\hat{\omega}_t^0 = \omega^0$. When the estimates are perfectly estimated, we will have both θ_t and LB_t perfectly estimated too. Hence, we have $F^0(\theta_t) - F^0(\hat{\omega}_t^0) = F^0(\hat{\omega}_t^0) - F^0(\text{LB}_t)$ such that $p_1 = p_2$. Thus, $\mathbb{E}[\hat{\omega}_{t+1}^0] = \omega^0$. Hence, once the parameter is correctly estimated, $\hat{f}_t^0$ is not expected to shift from f^0.

Case 2 (Underestimated): $\hat{\omega}_t^0 < \omega^0$. Under the unimodel distribution and single parameter assumption, since the arriving batch of data comes from the true distribution, $F^0(\text{LB}_t), F^0(\hat{\omega}_t^0), F^0(\theta_t)$ will be smaller than $\hat{F}^0(\text{LB}_t), \hat{F}^0(\hat{\omega}_t^0), \hat{F}^0(\theta_t)$, respectively. Moreover, since we have $\text{LB}_t \leq \hat{\omega}_t^0 \leq \theta_t \leq \hat{\omega}_t^1$, then $F^0(\theta_t) - F^0(\hat{\omega}_t^0) \geq F^0(\hat{\omega}_t^0) - F^0(\text{LB}_t)$ such that $p_2 \geq p_1$. Hence, 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 3 (Overestimated): $\omega^0 < \hat{\omega}_t^0$. Through similar analysis as Case 2 (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 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:

$$\mathbb{P}(\hat{\omega}_t^0 = \nu)\mathrm{d}\nu = \frac{(2m+1)!}{m!m!} \left(\frac{F^0(\nu) - F^0(\text{LB}_t)}{F^0(\theta_t) - F^0(\text{LB}_t)}\right)^m \left(\frac{F^0(\theta_t) - F^0(\nu)}{F^0(\theta_t) - F^0(\text{LB}_t)}\right)^m \frac{f^0(\nu)}{F^0(\theta_t) - F^0(\text{LB}_t)} \mathrm{d}\nu \tag{3}$$

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 $H(\nu) := \frac{F^0(\nu) - F^0(\text{LB}_t)}{F^0(\theta_t) - F^0(\text{LB}_t)}$; this is the CDF of the truncated F^0 distribution in $[\text{LB}_t, \theta_t]$. 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 [28] that for samples located in the range of $[\text{LB}_t, \theta_t]$, the sampling distribution of the median becomes asymptotically normal with mean $(\omega^0)'$ and variance $\frac{1}{4(2m+3)H((\omega^0)')}$, where $(\omega^0)'$ is the median, the truncated F^0 distribution in $[\text{LB}_t, \theta_t]$. If the sequence of $\{\hat{\omega}_t^0\}$ produced by our `active debiasing` algorithm converges, by Definition 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 $\hat{\omega}_{t+1}^0 \to (\omega^0)'$. By Definition 1, it must be that the median $(\omega^0)'$ 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 $\hat{y}^0_{t+1}$ is the next immediate update after $\hat{y}^0_t$ in the original sequence $\{\hat{\omega}^0_t\}$; that is, an underestimated $\hat{y}^0_t$ has been updated to a parameter that continues to be an underestimate. In this case, by Part (a), $\mathbb{E}[\hat{y}^0_{t+1}|\hat{y}^0_t] \geq \hat{y}^0_t$, and therefore, $\mathbb{E}[\Delta^y_{t+1}|\Delta^y_t] \leq \Delta^y_t$.

- Alternatively assume $\hat{y}^0_{t+1}$ is not obtained immediately from $\hat{y}^0_t$; that is, $\hat{y}^0_{t+1}$ has been obtained as a result of an update from an overestimated parameter. We note that now, $\hat{y}^0_{t+1} \geq \hat{y}^0_t$. This is because either no new estimates have been obtained between $\hat{y}^0_t$ and the true parameter ω^0 since the last time the parameter was underestimated, in which case, it must be that $\hat{y}^0_{t+1} = \hat{y}^0_t$. Otherwise, a new estimate in $[\hat{y}^0_t, \omega^0]$ has been obtained, in which case, again, $\mathbb{E}[\hat{y}^0_{t+1}|\hat{y}^0_t] \geq \hat{y}^0_t$. In either case, $\mathbb{E}[\Delta^y_{t+1}|\Delta^y_t] \leq \Delta^y_t$.

Therefore, by the Doobs Convergence theorem, the supermatingales $\{\Delta^y_t\}$ and $\{\Delta^z_t\}$ 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 $\{\hat{\omega}^0_t\}$ converges to ω. □

F Proof of Theorem 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. [2], and in learning under source and target distribution mismatches as proposed by Ben-David et al. [5]. We adapt these to our bounded exploration setting. In order to find our algorithm's error bound, we proceed through five steps. The first step is to view each individual update of the fair threshold classifier as a saddle point problem, which can be solved efficiently by the exponentiated gradient reduction method introduced in Agarwal et al. [2]. Second, based on the solution output from the reduction method, we find a bound on the classification error achievable based on data from the biased distributions. Thirdly, using results from Ben-David et al. [5], we bound the error on the target (unbiased) distribution when the algorithm is obtained from the the biased source domain. Then, we will evaluate the impact of exploration errors made by our debiasing algorithm. To simplify notation, we outline the error incurred on one group and at time $t = 0$; the effects of errors on other groups and in the subsequent time steps can be similarly obtained.

In more detail, our algorithm's error is made up of errors from four different sources, which we characterize step by step. Firstly, we have an *approximately* optimal classifer that is returned by the exponentiated gradient algorithm with suboptimiality level v (step 1). Secondly, we use samples to estimate the distributions and thus have empirical biases (step 2). Thirdly, since we start from biased distributions, there are errors due to the domain mismatches (step 3). Lastly, in order to debias, we explore by admitting samples that would otherwise be rejected, introducing additional errors (step 4).

Before proceeding, we outline our notation for different forms of data bias. First, note that there is a true underlying distribution for the population of agents to which the classifier is to be applied; we denote this by $\bar{D}$. Our focus in this work is on setting where there are different forms of statistical biases in the training data (e.g. distribution shifts or adaptive sampling biases); denote this statistically biased training data by $\tilde{D}$. Finally, even without distribution shifts or adaptive sampling biases, the classifier has access to a limited, empirically biased subset of this data; we denote the initial statistically and empirically biased data distribution by $\hat{D}$.

Accordingly, let $\hat{h}^*_{\theta_{g,0}}, \tilde{h}^*_{\theta_{g,0}}, \bar{h}^*_{\theta_g}$ be the optimal (fair and error minimizing) classifiers that would be obtained from an initial statistically and empirically biased dataset, only statistically biased dataset, and an unbiased dataset, respectively.

Step 1: Approximate solution errors. We can treat the problem of finding the initial fair classifier from the statistically and empirically biased training data as a saddle point problem. First, let $\tilde{err}(h_{\theta_{g,t=0}}) = \mathop{\mathbb{E}}_{(x_i,y_i,g_i)\sim\tilde{D}}\Big[\ell(h_{\theta_{g,t=0}}(x_i, g_i), y_i)\Big]$; this is the true error incurred by a classifier $h_{\theta_{g,t=0}}$ when training data comes from $\tilde{D}$, and is the objective function of the minimization problem. Additionally, we assume throughout that a fairness constraint $|\mathcal{C}(\theta_{a,t}, \theta_{b,t})| \leq \gamma$ has been imposed.

However, since we do not have the true $\tilde{D}$, and only have access to a limited, empirically biased subset of it $\hat{D}$, we will use the empirical estimates $\hat{err}(h_{\theta_{g,0}}), \hat{\mathcal{C}}(\theta_{a,0}, \theta_{b,0})$ and $\hat{\gamma}$ in the constrained optimization problem of finding the fair, loss-minimizing classifier.

To capture the fairness constraint, we will introduce Lagrangian multipliers $\lambda_j \geq 0$. This allows us to define the Lagrangian of the optimization problem:

$$\mathcal{L}(h_{\theta_{g,0}}, \lambda_j) = \hat{err}(h_{\theta_{g,0}}) + \lambda_1(\hat{\mathcal{C}}(\theta_{a,0}, \theta_{b,0}) - \hat{\gamma}) + \lambda_2(-\hat{\mathcal{C}}(\theta_{a,0}, \theta_{b,0}) - \hat{\gamma})$$

Following the rewriting procedures in Agarwal et al. [2] and using the exponentiated gradient algorithm, we can obtain a v-approximated solution $(\hat{h}_{\theta_{g,0}}, \hat{\lambda}_j)$; this is an approximately loss-minimizing fair classifier obtained based on an initial, empirically and statistically-biased training data, and the corresponding Lagrange multipliers of the fairness constraint.

Step 2: Empirical error bound on the initial biased distribution. 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. Let $n_{g,t}$ be the number of training samples arriving in round t from agents in group g. Initially, we have $n_{g,0} = b^0_{g,0} + b^1_{g,0}$. 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. [2], we can find that with probability at least $1 - 4\delta$ with $\delta > 0$:

$$\tilde{err}(\hat{h}_{\theta_{g,0}}) \leq \tilde{err}(\tilde{h}^*_{\theta_{g,0}}) + 2v + 4\mathcal{R}_{n_{g,0}}(\mathcal{H}) + \frac{4}{\sqrt{n_{g,0}}} + \sqrt{\frac{2\ln(2/\delta)}{n_{g,0}}} \tag{4}$$

In words, this provides a bound on the true error that will be incurred on statistically biased data when using the classifier obtained in step 1 (from statistically and empirically biased data).

Step 3: Bound of error on different distributions (domains). Next, we note that there is a mismatch between our current biased training data and the true underlying data distribution. We use results from domain adaptation to bound these errors.

To bound the error on different distributions, L^1 divergence would be a nature measure. However, it overestimates the bounds since it involves a supremum over all measurable sets. As discussed by Ben-David et al. [5], using classifier-induced divergence ($\mathcal{H}\Delta\mathcal{H}$-divergence) allows us to directly estimate the error of a source-trained classifier on the target domain by representing errors relative to other hypotheses in the hypothesis class.

Definition 2 ($\mathcal{H}\Delta\mathcal{H}$-divergence)**.** *For a hypothesis space $\mathcal{H}$, the symmetric difference hypothesis space $\mathcal{H}\Delta\mathcal{H}$ is the set of hypotheses:*

$$g \in \mathcal{H}\Delta\mathcal{H} \Leftrightarrow g(x) = h(x) \oplus h'(x) \quad \textit{for some } h, h' \in \mathcal{H}$$

where $\oplus$ is XOR function. In other words, every hypothesis g is the set of disagreement between two hypotheses in $\mathcal{H}$. The $\mathcal{H}\Delta\mathcal{H}$-distance is also given by

$$d_{\mathcal{H}\Delta\mathcal{H}}(D, D') = 2 \sup_{h,h' \in \mathcal{H}} \left| Pr_{x\sim D}[h(x) \neq h'(x)] - Pr_{x\sim D'}[h(x) \neq h'(x)] \right|$$

Let $\overline{err}(h)$ be the error made by a classifier h on unbiased data from the true underlying distribution. We bound this error below.

Lemma 1 (Follows from Theorem 2 of Ben-David et al. [5])**.** *Let $\mathcal{H}$ be a hypothesis space. If unlabeled samples are from $\tilde{D}_{g,0}$ and D_g respectively, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$:*

$$\overline{err}(\hat{h}_{\theta_{g,0}}) \leq \tilde{err}(\hat{h}_{\theta_{g,0}}) + \tfrac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,0}, D_g) + c(\tilde{D}_{g,0}, D_g) \tag{5}$$

where $\tilde{D}_{g,0}$ and D_g are the joint distribution of labels, and $c(\tilde{D}_{g,0}, D_g) = \min_h \overline{err}(h) + \tilde{err}(h)$.

Then, combining equation 4 and 5, we can obtain the following expression:

$$\begin{aligned}
\overline{err}(\hat{h}_{\theta_{g,0}}) &\leq \tilde{err}(\hat{h}_{\theta_{g,0}}) + \tfrac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,0}, D_g) + c(\tilde{D}_{g,0}, D_g) \\
&\leq \tilde{err}(\tilde{h}^*_{\theta_{g,0}}) + 2v + 4\mathcal{R}_{n_{g,0}}(\mathcal{H}) + \tfrac{4}{\sqrt{n_{g,0}}} + \sqrt{\tfrac{2\ln(2/\delta)}{n_{g,0}}} + \tfrac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,0}, D_g) + c(\tilde{D}_{g,0}, D_g) \\
&\leq \tilde{err}(\bar{h}^*_{\theta_{g,0}}) + 2v + 4\mathcal{R}_{n_{g,0}}(\mathcal{H}) + \tfrac{4}{\sqrt{n_{g,0}}} + \sqrt{\tfrac{2\ln(2/\delta)}{n_{g,0}}} + \tfrac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,0}, D_g) + c(\tilde{D}_{g,0}, D_g) \\
&\leq \overline{err}(\bar{h}^*_{\theta_{g,0}}) + 2v + 4\mathcal{R}_{n_{g,0}}(\mathcal{H}) + \tfrac{4}{\sqrt{n_{g,0}}} + \sqrt{\tfrac{2\ln(2/\delta)}{n_{g,0}}} + d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,0}, D_g) + 2c(\tilde{D}_{g,0}, D_g)
\end{aligned}$$

In words, this provides a bound on the true error that will be incurred on the unbiased data from the underlying population when using the classifier obtained in step 1 (from statistically and empirically biased data).

Step 4: Exploration errors. Lastly, in order to reduce the mismatches between the biased training data and the true underlying distribution, our algorithm incurs some exploration errors. Let $n'_{0,g,t}$ and $n'_{1,g,t}$ denote the number of samples from unqualified and qualified group that fall below the threshold $\theta_{g,t}$ in round t, respectively. Since in Steps 2 and 3 we already considered the classification errors due to empirical estimation and different distributions, we only consider the additional exploration error introduced with the goal of removing biases. Because of exploration, some samples from the qualified group that were rejected previously will now be accepted, which will allow the algorithm to make less errors. On the other hand, some samples from the unqualified group that would previously be rejected are now accepted, which will lead to an increase in the errors.

Denote ϵ_t as the exploration probability at round t. The exploration error consists of the errors made on the unqualified group, minus correct decisions made on the qualified group. In bounded exploration approach, we introduce a LB_t to limit the depth of exploration. Therefore, the number of samples that 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; this is given by:

$$N_{g,t} := \left(\frac{\hat{F}^0_{g,t}(\theta_t) - \hat{F}^0_{g,t}(\text{LB}_t)}{\hat{F}^0_{g,t}(\theta_t)} \epsilon_t n'_{0,g,t} - \frac{\hat{F}^1_{g,t}(\theta_t) - \hat{F}^1_{g,t}(\text{LB}_t)}{\hat{F}^1_{g,t}(\theta_t)} \epsilon_t n'_{1,g,t} \right)$$

Step 5: Errors made over m updates. We now state the error incurred by our algorithm over m rounds of updates. For a group g, combining the four identified sources of error over m updates, we have

$$\sum_{t=1}^{m} \overline{err}(\hat{h}_{\theta_{g,t}}) \leq \sum_{t=1}^{m} \left[\overline{err}(\bar{h}^*_{\theta_{g,t}}) + 2v + 4\mathcal{R}_{n_{g,t}}(\mathcal{H}) + \frac{4}{\sqrt{n_{g,t}}} + \sqrt{\frac{2\ln(2/\delta)}{n_{g,t}}} + N_{g,t} + d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,t}, D_g) + 2c(\tilde{D}_{g,t}, D_g) \right]$$

Therefore, the error bound for our algorithm over m updates and across two groups $g \in \{a, b\}$ is given by

$$\begin{aligned}
\text{Err.} &= \sum_{t=1}^{m} \left[\overline{err}(\hat{h}_{\theta_{a,t}}) + \overline{err}(\hat{h}_{\theta_{b,t}}) - \overline{err}(h^*_{\theta_{a,t}}) - \overline{err}(h^*_{\theta_{b,t}}) \right] \\
&\leq \sum_{g,t} \Bigg[\underbrace{2v}_{v\text{-approx.}} + \underbrace{4\mathcal{R}_{n_{g,t}}(\mathcal{H}) + \frac{4}{\sqrt{n_{g,t}}} + \sqrt{\frac{2\ln(2/\delta)}{n_{g,t}}}}_{\text{empirical estimation}} + \underbrace{N_{g,t}}_{\text{explor.}} + \underbrace{d_{\mathcal{H}\Delta\mathcal{H}}(\tilde{D}_{g,t}, D_g) + 2c(\tilde{D}_{g,t}, D_g)}_{\text{source-target distribution}} \Bigg]
\end{aligned}$$

□

From the expression above, we can see that the more samples we have, $n_{g,t}$ will increase. Hence, the empirical estimation error will decrease. Moreover, as the mismatch between $\tilde{D}_{g,t}$ and D_g is removed by our debiasing algorithm, the mismatch between the source and target domains will also decrease, decreasing the corresponding error term. In the meantime, when ϵ_t is set adaptively, the exploration probability becomes smaller as we remove the mismatch. Therefore, the exploration error $N_{g,t}$ will also decrease. Together, these mean that the terms in the summation above decrease as t increases.

G Proof of Proposition 1

Proof. We compare the speed of debiasing through $\mathbb{E}[|\hat{\omega}^y_t - \omega^y|]$. Given a fixed t, we say the algorithm for which this error is larger has a lower speed of debiasing. In words, the slower algorithm needs to wait for *more* arriving samples before it can reach the same parameter estimation error as a faster algorithm.

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 $\hat{f}_t^0$, which depends on the choice of LB_t in Definition 1, i.e.,

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

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

Now, consider the interval $[\text{LB}_t, \max^0]$, with $\max^0$ denoting the maximum of f^0. 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 $\text{LB}_t^F > \text{LB}_t^U$, this interval is narrower under the fairness constrained classifier, meaning that it takes more time to meet the batch size requirement under compared LB_t^U compared to LB_t^F. As detailed in the proof of Theorem 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 $\hat{f}_t^0$ is slower after the introduction of fairness constraints.

Similar arguments hold for updating $\hat{f}_t^1$, 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 $\hat{f}_t^1$ as well. □

H Larger figures and additional experiments

H.1 Larger figures and experiment details

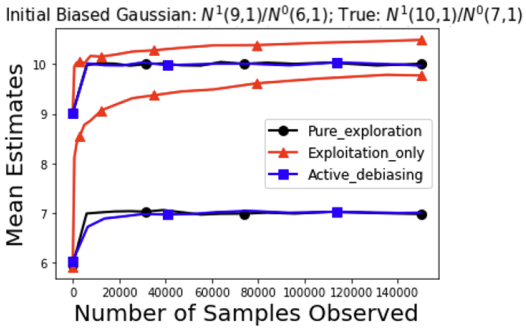


Figure 5: Rate of debiasing of our active debiasing algorithm vs the two baselines. Underlying feature distributions are Gaussian. We let f^1 be underestimated with $\hat{\omega}^1 = 9$ and true parameter $\omega^1 = 10$ (parameter debiasing shown in the top lines), and f^0 underestimated with $\hat{\omega}^0 = 6$ and true parameter $\omega^0 = 7$ (parameter debiasing shown in bottom lines).

H.2 Additional experiments: effects of depth of exploration

Figure 12 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' 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 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. 12(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. 12(b), the increased speed of debiasing means that the

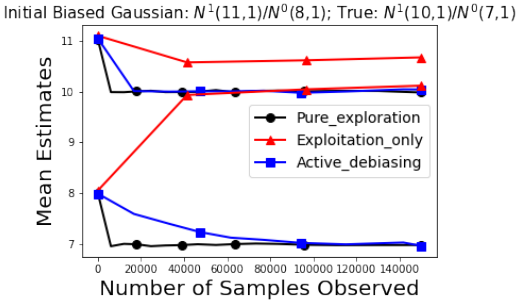


Figure 6: Rate of debiasing of our active debiasing algorithm vs the two baselines. Underlying feature distributions are Gaussian. We let f^1 be overestimated with $\hat{\omega}^1 = 11$ and true parameter $\omega^1 = 10$ (parameter debiasing shown in the top lines), and f^0 overestimated with $\hat{\omega}^0 = 8$ and true parameter $\omega^0 = 7$ (parameter debiasing shown in bottom lines).

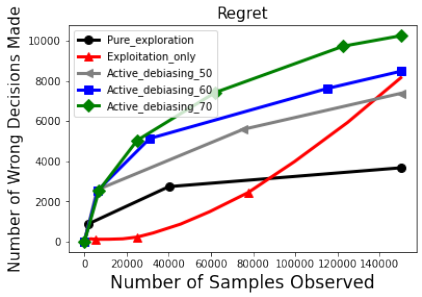


Figure 7: Regret, measured as the number of wrong decisions (sum of False Positives and False Negatives) compared to an oracle classifier which knows the unbiased underlying distributions. Underlying feature distributions are Gaussian. We let f^1 and f^0 be overestimated with $\hat{\omega}^1 = 9$ and $\hat{\omega}^0 = 6$, and their true parameters $\omega^1 = 10$ and $\omega^0 = 7$, respectively.

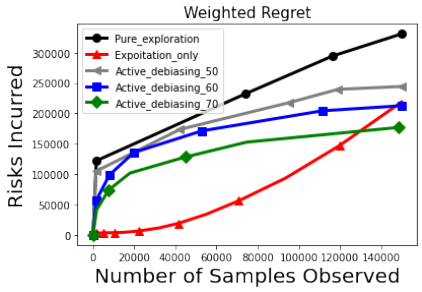
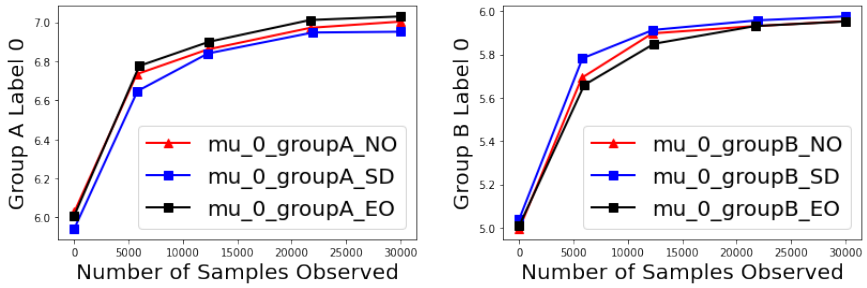


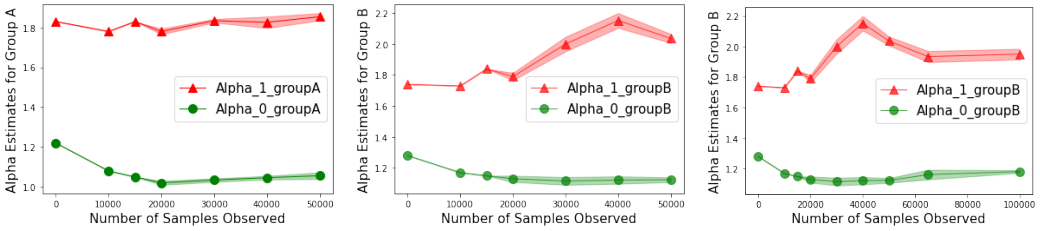
Figure 8: Weighted regret, measured as the *risk* of the wrong decisions made (a weighted sum of False Positives and False Negatives) compared to an oracle classifier which knows the unbiased underlying distributions. The weighted regret for each label 0 (resp. 1) sample is calculated by the exponential of the difference between its feature and the forth standard deviation above (resp. below) the mean. Underlying feature distributions are Gaussian. We let f^1 and f^0 be overestimated with $\hat{\omega}^1 = 9$ and $\hat{\omega}^0 = 6$, and their true parameters $\omega^1 = 10$ and $\omega^0 = 7$, respectively.



(a) Advantaged label 0. We assume We assume the assumed distribution is underestimated, with estimated parameter $\hat{\omega}_a^0 = 6$ and true parameter $\omega_a^0 = 7$.

(b) Disadvantaged label 1. We assume the assumed distribution is underestimated, with estimated parameter $\hat{\omega}_b^0 = 8$ and true parameter $\omega_b^0 = 9$.

Figure 9: `Active debiasing` under different fairness constraints. The underlying feature distributions are Gaussian, and the reference points are set to $\alpha^0 = 60$ and $\alpha^1 = 50$ for both groups. The exploration frequency $\{\epsilon_t\}$ is reduced with the fixed schedule of being subtracted by 0.1 after observing every 3000 samples.



(a) Debiasing G_a.

(b) Debiasing G_b.

(c) *Adult* dataset augmented with 60,000 synthetic samples.

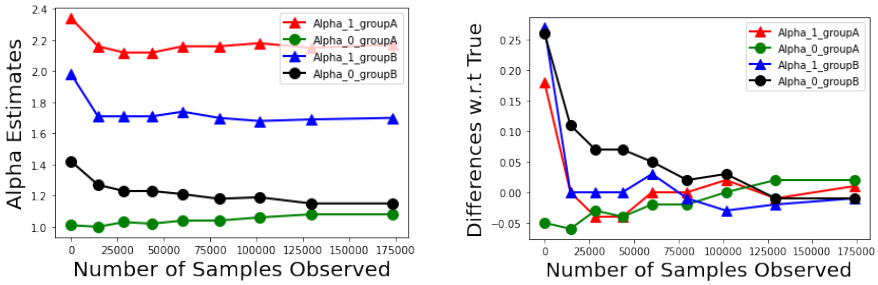
Figure 10: Illustration of the performance of `active debiasing` on the *Adult* dataset. The true underlying distributions were estimated to be Beta distributions with parameters Beta(1.94, 3.32) and Beta(1.13, 4.99) for group a (White) label 1 and 0, respectively, and Beta(1.97, 3.53) and Beta(1.19, 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

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).

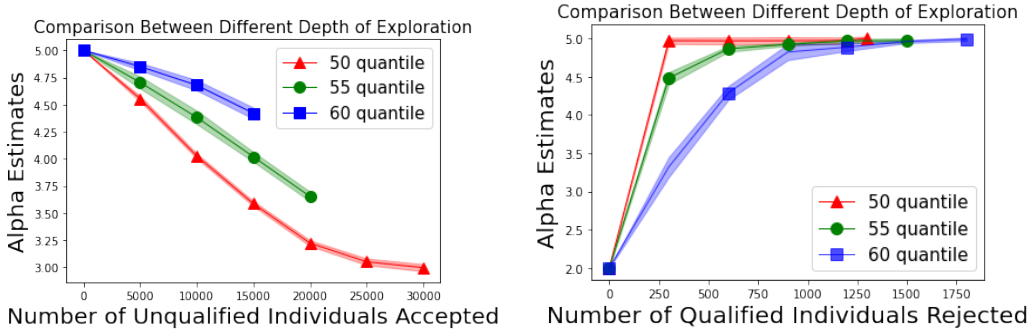
I Debiasing with two unknown parameters: a Gaussian distribution with two unknown parameters mean μ and variance σ^2

In this section, 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.



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

Figure 11: 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



(a) False positives (unqualified agents admitted) under each reference point (b) False negatives (qualified agents rejected) under each reference point

Figure 12: `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.

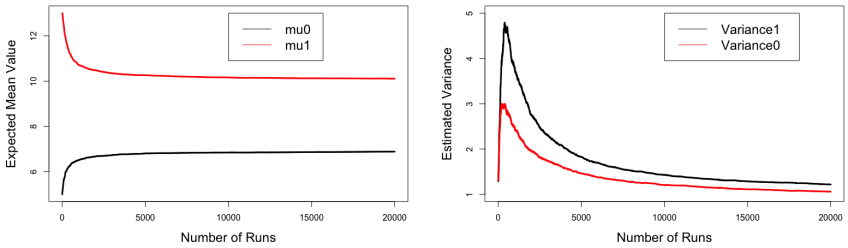
We follow our `active debiasing` algorithm, with a choice of medians as reference points (i.e., $\alpha^i = 50, \forall i$), and setting the thresholds UB (See Definition 3 below) and LB so that the reference points are the medians of the truncated distribution between the bounds and the classifier θ. For this experiment only, we set a UB similar to LB for simplicity. 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.

Definition 3. *At time t, the firm selects a upper bound UB_t such that*

$$UB_t = (\hat{F}_t^1)^{-1}(2\hat{F}_t^1(\hat{\omega}_t^1) - \hat{F}_t^1(LB_t)),$$

where LB_t is obtained from Definition 1, $\hat{F}_t^1$, $(\hat{F}_t^1)^{-1}$ are the cdf and inverse cdf of the estimated distribution $\hat{f}_t^1$, respectively, and $\hat{\omega}_t^1$ is (wlog) the α-th percentile of $\hat{f}_t^1$.

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 UB and LB in the data collection procedure. We therefore need to convert between the estimated statistics for the truncated distribution and those of the full distribution accordingly.



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

Figure 13: 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.

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 UB and LB, 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, $a = UB$ and $b = \theta$ for $i = 1$, and $a = \theta$ and $b = LB$ 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 13 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.