Bounding the Fairness and Accuracy of Classifiers from Population Statistics

Sivan Sabato and Elad Yom-Tov

ICML 2020
Supplementary Material

A. Algorithm for Finding the Accuracy of a Fair Classifier

Alg. 2 gives the full algorithm for lower-bounding the `error` of a classifier which is known to be fair. Its time complexity is linear in $|\mathcal{G}|$.

Algorithm 2 Lower-bounding the error of a classifier which is known to be fair

Input: `Inputs` $\equiv (\{w_g\}, \{\pi_g^y\}\{\hat{p}_g^y\})$ for a fair classifier
Output: A lower bound on the `error` of the classifier, or `UNFAIR` if the input is inconsistent with a fair classifier.

1: For $g \in \mathcal{G}^+$, $r_g \leftarrow 1 - \hat{p}_g^1/\pi_g^1$ and $q_g \leftarrow 1/\pi_g^1 - 1$.
2: Solve the following linear minimization problem:

$$\text{Minimize}_{\alpha_{\text{all}}^0, \alpha_{\text{all}}^1 \in [0,1]} \quad \texttt{error} \equiv \sum_{y \in \mathcal{Y}} \alpha_{\text{all}}^y \sum_{g \in \mathcal{G}} w_g \pi_g^y$$

such that

$$\forall g \in \mathcal{G}^+, \quad \alpha_{\text{all}}^1 = r_g + q_g \alpha_{\text{all}}^0$$
$$\forall g \in \mathcal{G} \setminus \mathcal{G}^+, y \in \mathcal{Y} \text{ s.t. } \pi_g^y = 1, \quad \alpha_{\text{all}}^y = 1 - \hat{p}_g^y.$$

3: **if** no solution exists **then**
4: **return** `UNFAIR`.
5: **else**
6: **return** `error` $= \sum_{y \in \mathcal{Y}} \alpha_{\text{all}}^y \sum_{g \in \mathcal{G}} w_g \pi_g^y$.
7: **end if**

B. Proof of Theorem 4.1

In this section, we prove Theorem 4.1. First, we show that minimizing Obj for a fixed $\bar{\alpha}$ can be done using a small finite number of possibilities for each variable in $\{\alpha_g^0\}_{g \in \mathcal{G}^+}$.

Lemma B.1. *Let* Obj *be as defined in Eq. (7) and let* Obj_2 *be as defined in Eq. (9). Then, for any* $\bar{\alpha} \in [0,1]^2$,

$$\text{Obj}_2(\bar{\alpha}) = \min_{\{\alpha_g^0 \in \text{dom}_g\}_{g \in \mathcal{G}^+}} \text{Obj}(\bar{\alpha}, \{\alpha_g^0\}_{g \in \mathcal{G}^+}).$$

Proof. Fix $\bar{\alpha} \in [0,1]^2$. We show that there exists an assignment that minimizes $\text{Obj}(\bar{\alpha}, \{\alpha_g^0\}_{g \in \mathcal{G}^+})$, which satisfies $\alpha_g^0 \in S_g(\bar{\alpha})$ for all $g \in \mathcal{G}^+$, where S_g is as defined in Theorem 4.1. From the definition of Obj_2, this suffices to prove the claim.

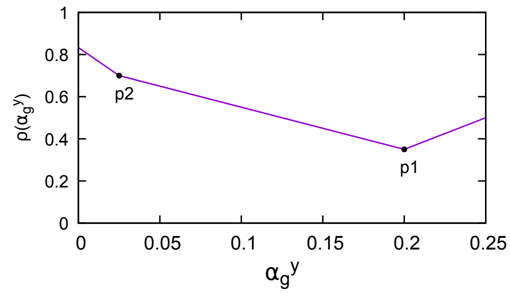

Figure 8. $\rho(\alpha_g^y)$ and the inflection points p_1, p_2 for $\alpha^0 = 0.2, \alpha^1 = 0.6, \pi_g^1 = 0.1, \hat{p}_g^1 = 0.1$.

From the definition of Obj, it suffices to show that for any fixed $\bar{\alpha}$, $\rho(\alpha_g^0) := \sum_y \pi_g^y \tau(\alpha^y, \alpha_g^y)$ is minimized by some $\alpha_g^y \in S_g(\bar{\alpha})$. Consider the function $b \mapsto \tau(a,b)$ for a fixed a. It is a convex combination of $\eta(a,b)$ and of b. Now, for a fixed $a \in [0,1]$, $b \mapsto \eta(a,b)$ is a convex piecewise-linear function: it is linear and decreasing over $[0,a]$ and linear and increasing over $[a,1]$, as can be verified from the definition of η (see Figure 1 (left) for illustration). Since by Eq. (5), α_g^1 is linear in α_g^0, it follows that $\rho(\alpha_g^0)$ is convex and piecewise-linear, with inflection points at $\alpha_g^0 = \alpha^0 =: p_1$, and at $\alpha_g^1 = \alpha^1$, that is at $\alpha_g^0 = (\alpha^1 - r_g)/q_g =: p_2$ (see Figure 8 for illustration). It follows that the minimizer of $\tau(\alpha_g^0)$ must be one of p_1, p_2 or the end points of dom_g. These are exactly the values in $S_g(\bar{\alpha})$ defined in Theorem 4.1. □

Next, we show that the set of possible minimizers for Obj_2 can be reduced to a finite family of one-dimensional solution sets. We note that since Obj_2 is only two-dimensional, a straight-forward approach to minimizing it might be to run a grid search over $[0,1]^2$ at some fixed resolution. However, when data is severely unbalanced, as in the case of cancer occurrence statistics, which we study in Section 5, the necessary grid size can be very large, and squaring this number can make the grid size infeasible. For instance, cancer occurrence statistics can be of the order 10^{-6}. A two-dimensional grid of this resolution would be of order 10^{12}, leading to a possibly impractical search size.

We start with the observation that each of the functions $s_g^i(\bar{\alpha})$ is continuous in α^0,α^1. Thus, Obj_2 can be redefined using continuous functions. Define

$$\forall g \in \mathcal{G}^+, i \in [4], \quad f_g^i(\bar{\alpha}) := \tag{11}$$
$$\begin{cases} \pi_g^0 \tau(\alpha^0, s_g^i(\bar{\alpha})) + \pi_g^1 \tau(\alpha^1, r_g + q_g s_g^i(\bar{\alpha})) & s_g^i(\bar{\alpha}) \in \text{dom}_g, \\ \infty & \text{otherwise.} \end{cases}$$

In addition, for $g \in \mathcal{G} \setminus \mathcal{G}^+$, define

$$f_g^1(\bar{\alpha}) := \sum_{y:\pi_g^y=1} w_g \tau(\alpha^y, 1-\hat{p}_g^y),$$

and $f_g^i := \infty$ for $i \in \{2,3,4\}$. Then, it easily follows that

$$\mathrm{Obj}_2(\bar{\alpha}) = \sum_{g\in\mathcal{G}^+} w_g \min_{i\in[4]} f_g^i(\bar{\alpha}).$$

The following lemma shows that each of the functions f_g^i is concave in α^1 on an appropriate split of its domain. A technical detail is that in some cases, $f_g^i(\bar{\alpha})$ might not be continuous in α^1 at $\alpha^1 = 0$ or at $\alpha^1 = 1$. This occurs because for $b \in \{0,1\}$, we have $1 = \lim_{a\to b} \eta(a,b) \neq \eta(b,b) = 0$. Thus, in some cases the concavity holds on a restricted domain, which does not include 0 or 1.

Lemma B.2. *Fix $\alpha^0 \in [0,1]$. For $g \in \mathcal{G}$, $i \in [4]$, consider the function $\alpha^1 \mapsto f_g^i(\alpha^0, \alpha^1)$ and the domain I_g^i on which it is finite. Define $\theta_g^i(\alpha^0)$ as follows:*

$$\theta_g^i(\alpha^0) = \begin{cases} \max\{r_g, 0\} & i=1, g\in\mathcal{G}^+; \\ \min\{r_g+q_g, 1\} & i=2, g\in\mathcal{G}^+; \\ r_g + q_g\alpha^0 & i\in\{3,4\}, g\in\mathcal{G}^+, \\ & \quad \alpha^0 \in \mathrm{dom}_g; \\ 1-\hat{p}_g^1 & i=1, g\in\mathcal{G}, \pi_g^1 = 1; \\ 0 & \textit{otherwise}. \end{cases}$$

Then $\alpha^1 \mapsto f_g^i(\alpha^0, \alpha^1)$ is concave on each of the intervals $(0, \theta_g^i(\alpha^0)] \cap I_g^i$ and $[\theta_g^i(\alpha^0), 1) \cap I_g^i$. Moreover, if $\theta_g^i(\alpha^0) \in (0,1)$, then this holds also for the closed intervals.

Proof. For given $g \in \mathcal{G}, i \in [4]$, define $v^0 := s_g^i(\bar{\alpha})$, and $v^1 := r_g + q_g v^0$. First, let $i \in [3]$, $g \in \mathcal{G}^+$, and assume $s_g^i(\alpha^0) \in \mathrm{dom}_g$. In this case, $f_g^i(\bar{\alpha}) = \pi_g^0 \tau(\alpha^0, v^0) + \pi_g^1 \tau(\alpha^1, v^1)$. Observe that s_g^1, s_g^2, s_g^3 are constant in α^1, thus v^0 does not depend on α^1, and the same holds for $\tau(\alpha^0, v^0)$. Therefore, from the definition of τ, the function $\alpha^1 \mapsto f_g^i(\bar{\alpha})$ is an affine function of $\eta(\alpha^1, v^1)$ (with non-negative coefficients). Note that v^1 does not depend on α^1. It is easy to verify that for any fixed $v^1 \in (0,1)$, $\alpha^1 \mapsto \eta(\alpha^1, v^1)$ is concave on the interval $[0, v^1]$ and on the interval $[v^1, 1]$ (see illustration in Figure 1 (right)). Therefore, $\alpha^1 \mapsto f_g^i(\bar{\alpha})$ is also concave on the same intervals. Substituting v^1 with its definition, we get $\theta_g^i(\alpha^0)$ as defined above. If $v^1 \in \{0,1\}$ then the function is concave on $(0, v^1]$ and on $[v^1, 1)$.

Next, consider $i = 4$, $g \in \mathcal{G}^+$. Note that we have $I_g^4 = [r_g, q_g + r_g] \cap [0,1]$, since $\alpha^1 \in I_g^4$ iff $v^0 \equiv s_g^4(\alpha^1) \in \mathrm{dom}_g$. In this case, we have $v^1 = \alpha^1$, thus $\eta(\alpha^1, v^1) = 0$ and so $\tau(\alpha^1, v^1)$ is linear in α^1. In addition, $\tau(\alpha^0, v^0)$ is an affine function of $\eta(\alpha^0, v^0) = \eta(\alpha^0, (\alpha^1 - r_g)/q_g)$. This function is linear in α^1 on each of the intervals $[0, r_g + q_g\alpha^0] \cap I_g^i$ and $[r_g + q_g\alpha^0, 1] \cap I_g^i$. Thus, $\alpha^1 \mapsto f_g^4(\bar{\alpha})$ is linear on each of these intervals.

Lastly, consider $g \in \mathcal{G} \setminus \mathcal{G}^+$. In this case, only $i = 1$ is finite. If $\pi_g^0 = 1$, then $\alpha^1 \mapsto f_g^1(\bar{\alpha})$ is constant. If $\pi_g^1 = 1$, then f_g^1 is an affine function with positive coefficients of $\eta(\alpha^1, 1 - \hat{p}_g^y)$. Similarly to the first case, this implies that $\alpha^1 \mapsto f_g^i(\bar{\alpha})$ is concave on $[0, 1-\hat{p}_g^1]$ and on $[1-\hat{p}_g^1, 1]$. □

From the concavity property proved in Lemma B.2, we can conclude that it is easy to minimize $\mathrm{Obj}_2(\bar{\alpha})$ over α^1 when α^0 is fixed, since there is only a small finite number of solutions. This is proved in the following lemma.

Lemma B.3. *Let $\alpha^0 \in [0,1]$, and define $\Theta(\alpha^0) := \{\theta_g^i(\alpha^0)\}_{g\in\mathcal{G}, i\in[4]} \cup \{0,1\}$. Then,*

$$\min_{\alpha^1\in[0,1]} \mathrm{Obj}_2(\bar{\alpha}) = \min_{\alpha^1\in\Theta(\alpha^0)} \mathrm{Obj}_2(\bar{\alpha}).$$

Proof. Fix α^0. Ordering the points in $\Theta(\alpha^0)$ in an ascending order and naming them $0 = p_1 < p_2 < \ldots < p_N = 1$, where $N = |\Theta(\alpha^0)|$, define the set of intervals $\mathcal{I} = \{[p_i, p_{i+1}] \setminus \{0,1\}\}_{i\in[N-1]}$. Observe that the intervals in $\mathcal{I}$ partition $(0,1)$ (with overlaps at end points). Moreover, for all $g \in \mathcal{G}^+, i \in [4]$, each of the intervals $I \in \mathcal{I}$ satisfies either $I \subseteq (0, \theta_g^i(\alpha^0)] \cap I_g^i$ or $I \subseteq [\theta_g^i(\alpha^0), 1) \cap I_g^i$. By Lemma B.2, it follows that for all $g \in \mathcal{G}^+, i \in [4]$, $\alpha^1 \mapsto \tilde{f}_g^i(\bar{\alpha})$ is concave on I.

Obj_2 is a conic combination of minima of concave functions on I, thus it is concave on I. Thus, on each $I \in \mathcal{I}$ which is a closed interval (that is, with endpoints other than 0 or 1), $\mathrm{Obj}_2(\bar{\alpha})$ is minimized by one of the endpoints of I. Now, consider I with an endpoint 0 or 1. For simplicity, assume $I = (0, p_2]$; the other case is symmetric. $\mathrm{Obj}_2(\bar{\alpha})$ is concave on I. Thus, it is either minimized on I at p_1 or satisfies $\lim_{\alpha_1\to 0} \mathrm{Obj}_2(\bar{\alpha}) \leq \inf_{\alpha^1\in I}(\mathrm{Obj}_2(\bar{\alpha}))$. In the latter case, there are two options: if $\alpha^1 \mapsto \mathrm{Obj}_2(\bar{\alpha})$ is continuous at 0, then $\mathrm{Obj}_2(\alpha^0, 0) = \lim_{\alpha_1\to 0} \mathrm{Obj}_2(\bar{\alpha})$, implying that the function is minimized on I by $\alpha^1 = 0$. If it is not continuous at 0, this can only happen if $r_g + q_g s_g^i(\bar{\alpha}) = 0$, which leads to $\eta(\alpha^1, r_g + q_g s_g^i(\bar{\alpha})) = 0$, whereas $\lim_{\alpha^1\to 0} \eta(\alpha^1, r_g + q_g s_g^i(\bar{\alpha})) = 1$. Thus, in this case, $\mathrm{Obj}_2(\alpha^0, 0) \leq \lim_{\alpha_1\to 0} \mathrm{Obj}_2(\bar{\alpha})$. It follows that in this case as well, Obj_2 is minimized on I by 0.

It follows that the minimizer of $\alpha^1 \mapsto \mathrm{Obj}_2(\bar{\alpha})$ must be one of the end points of $\Theta(\alpha^0)$, as claimed. □

Now, due to the complete symmetry of the problem definition between the two labels, Lemma B.3 implies also the symmetric property, that for a fixed $\alpha^1 \in [0,1]$, $\min_{\alpha^0\in[0,1]} \mathrm{Obj}_2(\bar{\alpha}) = \min_{\alpha^0\in\tilde{\Theta}(\alpha^1)} \mathrm{Obj}_2(\bar{\alpha})$, where $\tilde{\Theta}$ is symmetric to Θ and is obtained by switching the roles of

$y = 1$ and $y = 0$ in all definitions. By explicitly calculating the resulting values, we get that

$$\tilde{\Theta}(\alpha^1) = V_0 \cup \{(\alpha^1 - r_g)/q_g\}_{g \in \mathcal{G}^+},$$
$$\Theta(\alpha^0) = V_1 \cup \{r_g + q_g \alpha^0\}_{g \in \mathcal{G}^+},$$

Where V_0, V_1 are defined as in Theorem 4.1. Using the above results, we can now prove Theorem 4.1.

Proof of Theorem 4.1. From Lemma B.1, we have that $V^* = \min_{\bar{\alpha} \in [0,1]^2} \mathrm{Obj}_2(\bar{\alpha})$. Thus, we have left to show that there exists a minimizer for $\mathrm{Obj}_2(\bar{\alpha})$ in the set Sols defined in Theorem 4.1. From Lemma B.3, it follows that there exists a minimizing solution $(a^0, a^1) \in \mathrm{argmin}_{\bar{\alpha} \in [0,1]^2} \mathrm{Obj}_2(\bar{\alpha})$ such that $a^1 \in \Theta(a^0)$. The symmetric counterpart of Lemma B.3 implies that there exists some $b^0 \in \tilde{\Theta}(a^1)$ such that $(b^0, a^1) \in \mathrm{argmin}_{\bar{\alpha} \in [0,1]^2} \mathrm{Obj}_2(\bar{\alpha})$. Thus, at least one of the following options must hold:

- $(b^0, a^1) \in V_0 \times V_1$.
- $a^1 \notin V_1$; this implies that $a^1 = r_g + q_g a^0$ for some $g \in \mathcal{G}^+$.
- $b^0 \notin V_0$; this implies that $b^0 = (a^1 - r_g)/q_g$ for some $g \in \mathcal{G}^+$.

In the first case, there exists a minimizer in $V_0 \times V_1$. In the second and third case, there exists a minimizer in the set $\{(v, r_g + q_g v) \mid v \in [0,1], g \in \mathcal{G}^+\}$. Thus, in all cases there is a minimizer in the set Sols, as required. □