On Statistical Discrimination as a Failure of Social Learning: A Multi-Armed Bandit Approach*

Junpei Komiyama[1] and **Shunya Noda**[2]

[1]Leonard N. Stern School of Business, New York University

[2]Vancouver School of Economics, University of British Columbia

{junpei.komiyama, shunya.noda}@gmail.com

Abstract

We analyze statistical discrimination using a multi-armed bandit model where myopic firms face candidate workers arriving with heterogeneous observable characteristics. The association between the worker's skill and characteristics is unknown ex ante; thus, firms need to learn it. In such an environment, laissez-faire may result in a highly unfair and inefficient outcome—myopic firms are reluctant to hire minority workers because the lack of data about minority workers prevents accurate estimation of their performance. Consequently, minority groups could be *perpetually underestimated*—they are never hired, and therefore, data about them is never accumulated. We proved that this problem becomes more serious when the population ratio is imbalanced, as is the case in many extant discrimination problems. We consider two affirmative-action policies for solving this dilemma: One is a subsidy rule that is based on the popular upper confidence bound algorithm, and another is the Rooney Rule, which requires firms to interview at least one minority worker for each hiring opportunity. Our results indicate temporary affirmative actions are effective for statistical discrimination caused by data insufficiency.

1 Introduction

Statistical discrimination refers to discrimination against minority people, taken by fully rational and non-prejudiced agents. In contrast to taste-based discrimination, which regards agents' preferences (e.g., racism, sexism) as the primary source of discrimination, the model of statistical discrimination does not assume any preferences towards specific groups. Previous studies have shown that even in the absence of prejudice, discrimination could occur persistently because of various reasons, such as the discouragement of human capital investment [Arrow, 1973; Foster and Vohra, 1992; Coate and Loury, 1993; Moro and Norman, 2004], information friction [Phelps, 1972; Cornell and Welch, 1996], and search friction [Mailath *et al.*, 2000]. The literature has proposed a variety of affirmative-action policies to solve statistical discrimination, and many of them are being implemented in practice.

The contribution of this paper is to articulate a new channel of statistical discrimination—underestimation of minority workers that appears as a consequence of social learning. Most of the extant literature focuses on behaviors of rational agents under an equilibrium where agents have a correct belief about the relationship between observable characteristics and unobservable skills. However, several empirical studies have shown that real-world people often have a biased belief towards minority groups. The aim of this study is to endogenize the evolution of the biased belief and analyze its consequence. In our model, (i) all firms (decision makers) are fully rational and non-prejudiced (i.e., attempt to hire the most productive worker), and (ii) all workers are ex ante symmetric. We show that, even in such an environment, a biased belief could be generated endogenously and persist in the long run.

We develop a *multi-armed bandit model* of social learning, in which many myopic and short-lived firms sequentially make hiring decisions. In each round, a firm faces multiple candidate workers. Each firm wants to hire only one person. Each firm's utility is determined by the hired worker's skill, which cannot be observed directly until employment. However, as in the standard statistical discrimination model, each worker also has an observable characteristic that is associated with the worker's hidden skill. In the beginning, no one knows the precise way to interpret the worker's observable characteristic for predicting that skill. Hence, firms first need to learn the relationship between the characteristic and skill, and then apply the statistical model to evaluate the predicted skill of workers. We assume that, firms submit all the information about their hiring cases to a public database, and therefore, each firm can observe all the past hiring cases (the characteristics and skills of all the workers *actually hired* in the past).

Each worker belongs to a group that represents the

*The full paper is available at: https://arxiv.org/abs/2010.01079.

worker's gender, race, and ethnicity. We assume that the characteristics of workers who belong to different groups should be interpreted differently. This assumption is realistic. First, previous studies have revealed that the underrepresented groups receive unfairly low evaluations in many places. When the observable characteristic is an evaluation provided by an outside rater, then the characteristic information itself could be biased because of the prejudice of the rater. Second, evaluations may also reflect differences in the culture, living environment, and social system. For example, firms must be familiar with the custom of writing recommendation letters to interpret letters correctly. Hence, the observable characteristics (curriculum vitae, exam score, grading report, recommendation letter, teaching evaluation, etc.) may provide very different implications even when their appearances are similar. If firms are impartial and aware of these biases, they should adjust the way they interpret the characteristics, by applying different statistical models for different groups.

The lack of data results in inaccurate prediction of minority workers' skills, and the inaccuracy discourages firms from hiring the minorities. Many workers apply for each job opening. To get hired, a worker must have the highest predicted skill. Once the minority group is underestimated, it is difficult for a minority worker to appear to be the best candidate worker—even if the true skill is the highest, the firm will not be convinced. Underestimation rarely happens once society acquires a sufficiently rich data set. However, in the very beginning of the hiring process, the minority group is underestimated due to bad realizations of the unpredictable component.

The structure described above causes *perpetual underestimation*. Firms tend to hire majority workers because of the imbalance of data richness. However, as long as firms only hire majority workers, society cannot learn about the minority group; thus, the imbalance remains even in the long run. Here, the minority group is perpetually underestimated: the lack of data prevents hiring, and therefore, minority workers are never hired. We prove that, perpetual underestimation may happen with a significantly large probability in our model. Importantly, perpetual underestimation becomes more frequent when the population ratio is imbalanced, as observed in many real-world discrimination problems.

The social discrimination triggered by perpetual underestimation is not only unfair but also socially inefficient. From the welfare perspective, if the time horizon (the total number of hiring opportunities) is sufficiently long, then it is not very costly to "experimenting" with a small fraction of minority workers for learning. However, because firms are selfish and myopic, they are not willing to bear the cost of experiments on their own. Here, *laissez-faire* results in the underprovision of a public good—the information about minority groups. By enforcing or incentivizing early movers (firms) to review the minority groups, late movers can refer to a more useful data set of hiring cases, leading to improvement of social welfare. Note that the policy intervention need not be persistent: once sufficiently rich data are collected, the government can terminate the affirmative action and return to laissez-faire.

2 Model

We develop a linear contextual bandit problem with myopic agents (firms). We consider a situation where N firms (indexed by $n = 1, \dots, N$) sequentially hire one worker for each. In each round n, a set of workers $I(n)$ arrives. Each worker $i \in I(n)$ takes no action, and firm n selects one worker $\iota(n) \in I(n)$. Both firms and workers are short-lived. Once round n is finished, firm n's payoff is finalized, all the workers not hired leave the market. Due to page limitation, all the proofs are omitted. See our full paper of the details.

Each worker i belongs to a group $g \in G = \{1, 2\}$.[1] We typically denote group 1 as a majority (dominant) group, and group 2 as a minority (discriminated) group. We assume that the population ratio is fixed: for every round n, the number of arrived workers who belong to group g is $K_g \in \mathbb{N}$ and $K = \sum_{g \in G} K_g$. Slightly abusing the notation, we denote the group worker i belongs to by $g(i)$. Each worker $i \in I$ also has an observable characteristic $\boldsymbol{x}_i \in \mathbb{R}^d$, where $d \in \mathbb{N}$ is its dimension. Finally, each worker i also has a skill $y_i \in \mathbb{R}$, which is not observable until worker i is hired. The characteristics and skills are random variables.

Since each firm's payoff is equal to the hired worker's skill y_i (plus the subsidy assigned to worker i as an affirmative action, if any), firms want to predict the skill y_i based on the characteristics $\boldsymbol{x}_i$. We assume that the characteristics and skills are associated in the following way:

$$y_i = \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)} + \epsilon_i,$$

where $\boldsymbol{\theta}_g \in \mathbb{R}^d$ is a *coefficient parameter*, and $\epsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ i.i.d. is an unpredictable error term. We assume $||\boldsymbol{\theta}_g|| \leq S$ for some $S \in \mathbb{R}_+$, where $||\cdot||$ is the standard L2-norm. Since ϵ_i is unpredictable,

$$q_i \coloneqq \boldsymbol{x}_i' \boldsymbol{\theta}_{g(i)}$$

is the best predictor of worker i's skill y_i.

The coefficient parameters $(\boldsymbol{\theta}_g)_{g \in G}$ are unknown in the beginning. Hence, unless firms share the information about the past hiring cases, firms are unable to predict each worker's skill y_i. We assume that all firms share the information about hiring cases. Accordingly, each firm can observe the characteristics and true skill of the workers hired thus far. Firms predict the arriving workers' skill using this data and make hiring decisions.

We assume that firms are not Bayesian but *frequentist*. Hence, firms have no prior distribution but they estimate the true parameter $\boldsymbol{\theta}$ using the available data set.

We assume that each firm predicts the skill by using *ridge regression* (also known as regularized least square)

[1] Some of our results do not rely on this two-group assumption. See our full paper for the detail.

to stabilize the small-sample inference. Let $N_g(n)$ be the number of rounds at which group-g workers are hired by round n. Let $\boldsymbol{X}_g(n) \in \mathbb{R}^{N_g(n)\times d}$ be a matrix that lists the characteristics of group-g workers hired by round n: each row of $\boldsymbol{X}_g(n)$ corresponds to $\{\boldsymbol{x}_{\iota(n')} : \iota(n') = g\}_{n'=1}^{n-1}$. Likewise, let $Y_g(n) \in \mathbb{R}^{N_g(n)}$ be a vector that lists the skills of group-g workers hired by round n: each element of $Y_g(n)$ corresponds to $\{y_{\iota(n')} : \iota(n') = g\}_{n'=1}^{n-1}$. We define $\boldsymbol{V}_g(n) \coloneqq (\boldsymbol{X}_g(n))'\boldsymbol{X}_g(n)$. For a parameter $\lambda > 0$, we define $\bar{\boldsymbol{V}}_g(n) = \boldsymbol{V}_g(n) + \lambda \boldsymbol{I}_d$, where $\boldsymbol{I}_d$ denotes the $d\times d$ identity matrix. Firm n estimates the parameter as follows:

$$\hat{\boldsymbol{\theta}}_g(n) \coloneqq (\bar{\boldsymbol{V}}_g(n))^{-1}(\boldsymbol{X}_g(n))'Y_g(n).$$

In the beginning of the game, the government commits to a *subsidy rule* $s_i(n,\cdot) : H \to \mathbb{R}_+$, which maps a history to a subsidy amount. Hence, once a history $h(n)$ is specified, firm n can identify the subsidy assigned to each worker $i \in I(n)$. Firm n attempts to maximize

$$\mathbb{E}_n\left[y_i + s_i(n)\right] = \hat{q}_i(n) + s_i(n).$$

In the beginning of the game, the government commits to a *subsidy rule* $s_i(n,\cdot) : H \to \mathbb{R}_+$, which maps a history to a subsidy amount. Hence, once a history $h(n)$ is specified, firm n can identify the subsidy assigned to each worker $i \in I(n)$. Firm n attempts to maximize

$$\mathbb{E}\left[y_i + s_i(n; h(n))\middle|\, h(n)\right] = \hat{q}_i(n; h(n)) + s_i(n; h(n)).$$

We measure social welfare by the smallness of *regret*, which is the standard measure to evaluate the performance of algorithms in multi-armed bandit models. The regret is defined as follows:

$$\mathrm{Reg}(N) \coloneqq \sum_{n=1}^{N}\left\{\max_{i\in I(n)} q_i - q_{\iota(n)}\right\}.$$

Since ϵ_i is unpredictable, it is natural to evaluate the performance of the algorithm (or the equilibrium consequence of the policy intervention) by checking the value of predictors q_i. If the parameter $(\boldsymbol{\theta}_g)_{g\in G}$ were known, each firm could easily calculate q_i for each worker i and choose $\iota(n) = \arg\max_{i\in I(n)} q_i$. In this case, the regret would become zero. However, since $(\boldsymbol{\theta}_g)_{g\in G}$ is unknown, it is too demanding to aim at zero regret. The goal of the policy design is to set up a mechanism that minimizes the expected regret $\mathbb{E}[\mathrm{Reg}(N)]$, where expectation is taken on random draw on the workers.[2] This aim is equivalent to maximization of the sum of the skills of the hired workers.

Some of the policies we study incentivize exploration by subsidization. The total budget required by a subsidy rule is also an important policy concern. The total amount of the subsidy is given by

$$\mathrm{Sub}(N) \coloneqq \sum_{n=1}^{N} s_{\iota(n)}(n).$$

[2]All the mechanisms proposed in this paper is deterministic. Hence, there is no algorithmic randomness.

3 Results

Since our focus is on *endogenous* generation of statistical discrimination, we assume that the groups have no fundamental difference; i.e., $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}$ and $\boldsymbol{x}_i \sim \mathcal{N}(\boldsymbol{\mu}_x, \sigma_x \boldsymbol{I}_d)$ i.i.d. (thus, the distribution of $\boldsymbol{x}_i$ does not depend on the group worker i belongs to).[3] While the *true* parameter $\boldsymbol{\theta}$ is common, we do not assume that firms are aware of this fact. Accordingly, $\hat{\boldsymbol{\theta}}_1(n)$ and $\hat{\boldsymbol{\theta}}_2(n)$ are typically different.

3.1 Laissez faire

First, we analyze the equilibrium consequence of *laissez faire*, which makes no policy intervention to the hiring procedure. Under laissez faire, each firm n myopically hires a worker whose predicted skill $\hat{q}_i(n)$ is the highest among all arriving workers $I(n)$. This decision rule coincides with the *greedy algorithm*, which has widely been studied in the multi-armed bandit literature.

The following theorem states that laissez faire achieves no regret in a long run: When the groups are ex ante symmetric and the population ratio is equal, the expected regret of laissez faire is shown to be $\tilde{O}(\sqrt{N})$

Theorem 1 (Sublinear Regret with Balanced Population)**.** Suppose that $K_g = 1$ for $g = 1, 2$. Then, the expected regret is bounded as

$$\mathbb{E}[\mathrm{Reg}^{\mathrm{LF}}(N)] \le C_{\mathrm{bal}}\sqrt{N}$$

where C_{bal} is a $\tilde{O}(1)$ factor that depends on model parameters. Here, $\tilde{O}(1)$ is a Landau notation that ignores polylogarithmic factors. Letting $\mu_x = ||\boldsymbol{\mu}_x||$, the factor C_{bal} is inverse proportional to $\Phi^c(\mu_x/\sigma_x)$, which is approximately scales as $\exp(-(\mu_x/\sigma_x)^2/2)$.

Theorem 2 shows that statistical discrimination is resolved spontaneously when the variation of candidate is large. The variation in characteristics naturally incentivizes selfish agents to explore the underestimated group, and therefore, with some additional conditions, we can bound the probability of perpetual underestimation.

Theorem 1 crucially relies on one unrealistic assumption—the balanced population ratio. In many real-world problems, the population ratio is imbalanced. The dominant group is often the majority of the population, and the discriminated is minority. Our next theorem shows that laissez faire frequently results in perpetual underestimation if the population is imbalanced.

Theorem 2 (Large Regret with Imbalanced Population)**.** Suppose that $K_2 = 1$ and $d = 1$. Let $K_1 > \log_2 N$. Then, under the laissez-faire decision rule,

$$\mathbb{E}\left[\mathrm{Reg}(N)\right] \ge C_{\mathrm{imb}} N = \tilde{\Omega}(N),$$

where $C_{\mathrm{imb}} = \tilde{\Theta}(1)$.

[3]Some of our results do not depend on this assumption. See the full paper for the detail.

Theorem 2 indicates that we should not be too optimistic about the consequence of laissez faire. The imbalance in the population ratio naturally favors the majority group by helping society to collect a richer data set of them, leading to statistical discrimination. This insight applies to many real-world problems because imbalanced population is a commonplace.

3.2 The UCB Mechanism

This section proposes a subsidy rule to resolve such a perpetual underestimation. We use the idea of the *upper confidence bound* (UCB) algorithm, which is widely used in the literature of bandit problem.

The UCB decision rule hires a worker who has the highest *UCB index* $\hat{q}_i(n)$, which is defined as follows:

$$\tilde{q}_i(n) \coloneqq \max_{\bar{\boldsymbol{\theta}}_{g(i)} \in \mathcal{C}_{g(i)}(n)} \boldsymbol{x}_i' \bar{\boldsymbol{\theta}}_{g(i)}.$$

Here, $\mathcal{C}_g(n)$ is the *confidence interval* proposed by Abbasi-Yadkori *et al.* [2011]. Just as the standard confidence interval, $\mathcal{C}_g(n)$ shrinks as firm n has a richer set of the data about group g. See our full paper for the detail of $\mathcal{C}_g(n)$.

This decision rule can be implemented by assigning an appropriate amount of subsidy to each candidate workers. For example, we can subsidize $s(n) = \tilde{q}_i(n) - \hat{q}_i(n)$ to align firm n's incentive with the UCB index.

The following theorem indicates that, the UCB mechanism achieves $\tilde{O}(\sqrt{N})$ regret with the subsidy of $\tilde{O}(\sqrt{N})$.

Theorem 3. (Sublinear Regret of UCB) Let $\lambda \geq \max(1, L^2)$. Then, by choosing sufficiently small δ, the regret under the UCB decision rule is bounded as

$$\mathbb{E}[\mathrm{Reg}(N)] \leq C_{\mathrm{ucb}} \sqrt{N},$$

where C_{ucb} is a $\tilde{O}(1)$ factor to N that depends on model parameters. Furthermore, the subsidy required for implementing the UCB decision rule is bounded as

$$\mathbb{E}[\mathrm{Sub}(N)] \leq C_{\mathrm{ucb}} \sqrt{N}.$$

3.3 The Hybrid Mechanism

A practical complication of implementing UCB because they assign subsidies forever: Although the confidence interval $\mathcal{C}_g(n)$ shrinks as n grows large, it does not degenerate to a singleton for any finite n. Accordingly, even for a large n, there remains a gap between expected skill $\hat{q}_i(n)$ and the UCB index $\tilde{q}_i(n)$ (though small in size). This feature is not desirable.

To overcome these limitations of the UCB mechanism, we propose the *hybrid mechanism*, which starts with the UCB mechanism but turns to laissez faire by terminating the subsidy at some point. We terminate the UCB-phase once the amount of data of the minority group is enough to induce spontaneous exploration. We prove that, our hybrid mechanism has $\tilde{O}(\sqrt{N})$ regret (as the UCB mechanism does), and its expected total subsidy amount is $\tilde{O}(1)$ (as opposed to $\tilde{O}(\sqrt{N})$ subsidy of UCB). We also confirmed that the hybrid mechanism needs a smaller budget than the UCB mechanism through a numerical simulation.

3.4 The Rooney Rule

Although the UCB-based subsidy rule is a powerful policy intervention to resolve statistical discrimination, the subsidy rule is sometimes difficult to implement in practice. This section articulates advantages and disadvantage of the *Rooney Rule*, which requires each firm to invite at least one candidate of each group to an on-sight interview. The Rooney Rule is relatively easy to implement because it requires neither the subsidy nor hard hiring quota, and therefore, it is adopted in various situations.

In the full paper, we analyze the performance of the Rooney Rule. Even after extending our model to a two-stage model (where a firm can observe an additional signal about the skill of *finalists*), laissez faire falls in perpetual underestimation with a significant probability, and therefore, it suffers from linear regret.

The Rooney Rule can solve this problem. If the additional signal is informative enough, even when a minority worker is underestimated, he or she may overturn the situation at the on-sight interview. This promotes the data accumulation about minority workers and successfully prevent perpetual underestimation. While we find that the Rooney Rule also has linear regret because of underexploitation, we also show that this disadvantage can be mitigated by applying the Rooney Rule *temporarily*.

4 Summary of Technical Contributions

We have also made several technical contributions on the top of the literature of multi-armed bandits Lai and Robbins [1985]; Auer *et al.* [2002]; Abbasi-Yadkori *et al.* [2011]; Kannan *et al.* [2017, 2018]: (i) the analysis of the greedy algorithm under heterogeneous populations, (ii) proposal of the hybrid approach, and (iii) analysis of the Rooney Rule. Details are described in Section 8 of the full paper.

5 Conclusion

We studied statistical discrimination by using a contextual multi-armed bandit model. Our dynamic model articulates that statistical discrimination can be caused by failure of social learning. The failure is more likely to occur when the population is imbalanced. We analyzed two policy interventions for mitigating statistical discrimination: The subsidy rules and the Rooney Rule. Our analyses provide a consistent practical policy implication: Affirmative actions are useful for resolving statistical discrimination caused by the data insufficiency, but it should be implemented as a temporal policy, rather than a permanent one.

References

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.

Kenneth Arrow. The theory of discrimination. In Orley Ashenfelter and Albert Rees, editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, 1973.

Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time Analysis of the Multiarmed Bandit Problem. *Machine Learning*, 47(2):235–256, May 2002.

Stephen Coate and Glenn C Loury. Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, pages 1220–1240, 1993.

Bradford Cornell and Ivo Welch. Culture, information, and screening discrimination. *Journal of Political Economy*, 104(3):542–571, 1996.

D. Foster and R. Vohra. An economic argument for affirmative action. *Rationality and Society*, 4:176 – 188, 1992.

Sampath Kannan, Michael Kearns, Jamie Morgenstern, Mallesh Pai, Aaron Roth, Rakesh Vohra, and Zhiwei Steven Wu. Fairness incentives for myopic agents. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, pages 369–386, 2017.

Sampath Kannan, Jamie H Morgenstern, Aaron Roth, Bo Waggoner, and Zhiwei Steven Wu. A smoothed analysis of the greedy algorithm for the linear contextual bandit problem. In *Advances in Neural Information Processing Systems*, pages 2227–2236, 2018.

T.L Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4 – 22, 1985.

George J Mailath, Larry Samuelson, and Avner Shaked. Endogenous inequality in integrated labor markets with two-sided search. *American Economic Review*, 90(1):46–72, 2000.

Andrea Moro and Peter Norman. A general equilibrium model of statistical discrimination. *Journal of Economic Theory*, 114(1):1–30, January 2004.

Edmund S Phelps. The statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661, 1972.