A MARKOV CHAIN CONVERGENCE THEOREM FOR GENERAL STATE SPACES

In this section, we present the Markov convergence theorem for general state spaces, as well as the conditions to satisfy the conditions of the theorem. We mainly follow the references of Roberts & Rosenthal (2004); Meyn & Tweedie (2012a); Asmussen & Glynn (2010); Scheutzow & Schindler (2021).

Notation A.1. *The following notations will be used.*

1. *$\mathcal{X}$ denotes a standard measurable space (aka standard Borel space), like $\mathcal{X} = \mathbb{R}^D$ or $\mathcal{X} = \mathbb{N}$, etc.*
2. *We use $\mathcal{B}_{\mathcal{X}}$ to denote the σ-algebra of (Borel subsets of) $\mathcal{X}$.*
3. *$T : \mathcal{X} \dashrightarrow \mathcal{X}$ denotes a Markov kernel (aka transition probability) from $\mathcal{X}$ to $\mathcal{X}$, i.e. formally a measurable map $T : \mathcal{X} \to \mathcal{P}(\mathcal{X})$ from $\mathcal{X}$ to the space of probability measures over $\mathcal{X}$.*
4. *For a point $x \in \mathcal{X}$ and measurable set $A \in \mathcal{B}_{\mathcal{X}}$ we write T similar to a conditonal probability distribution:*
$$\begin{aligned} T(A|x) &:= T_x(A) \\ &:= \textit{ probability of } T \textit{ hitting } A \\ &\textit{when starting from point } x. \end{aligned} \tag{7}$$
5. *We define the Markov kernel $T^0 : \mathcal{X} \dashrightarrow \mathcal{X}$ via: $T^0(A|x) := 1_A(x)$.*
6. *We inductively define the Markov kernels $T^n : \mathcal{X} \dashrightarrow \mathcal{X}$ for $n \in \mathbb{N}_1$ via:*
$$\begin{aligned} T^n(A|x) &:= \int_{\mathcal{X}} T(A|y)\, T^{n-1}(dy|x) \\ &= \overbrace{(T \circ T \circ \cdots \circ T \circ T)}^{n\text{-times}}(A|x). \end{aligned} \tag{8}$$
Note that: $T^1 = T$.
7. *As the sample spaces we consider the product space:*
$$\Omega := \prod_{n \in \mathbb{N}_1} \mathcal{X}. \tag{9}$$
8. *For $n \in \mathbb{N}_1$ we have the canonical projections:*
$$\begin{aligned} X_n : \Omega &\to \mathcal{X}, \\ \omega = (x_n)_{n \in \mathbb{N}_1} &\mapsto x_n =: X_n(\omega). \end{aligned} \tag{10}$$
9. *We use $P_x := T_x^{\otimes \mathbb{N}_1}$ to denote the probability measure on Ω of the homogeneous Markov chain induced by T that starts at $X_0 = x$. Note that for $n \in \mathbb{N}_1$ the marginal distribution is given by:*
$$P_x(X_n \in A) = T^n(A|x). \tag{11}$$
10. *We abbreviate the tuple: $\mathbf{X} := (X_n)_{n \in \mathbb{N}_1}$. Note that $\mathbf{X}$ is a (homogeneous) Markov chain that starts at $X_0 = x$ under the probability distribution P_x. We will thus also refer to $\mathbf{X}$ as the (homogeneous) Markov chain corresponding to T.*
11. *We abbreviate the probability of the Markov chain of ever hitting $A \in \mathcal{B}_{\mathcal{X}}$ when starting from $x \in \mathcal{X}$ as:*
$$L(A|x) := P_x\left(\bigcup_{n \in \mathbb{N}_1} \{X_n \in A\} \right). \tag{12}$$

12. *We abbreviate the probability of the Markov chain hitting $A \in \mathcal{B}_{\mathcal{X}}$ infinitely often when starting from $x \in \mathcal{X}$ as:*

$$Q(A|x) := P_x\left(\{X_n \in A \text{ for infinitely many } n \in \mathbb{N}_1\}\right). \tag{13}$$

13. *We abbreviate the expected number of times the Markov chain hits $A \in \mathcal{B}_{\mathcal{X}}$ when starting from $x \in \mathcal{X}$ as:*

$$\begin{aligned} U(A|x) &:= \sum_{n \in \mathbb{N}_1} T^n(A|x) = \mathbb{E}_x[\eta_A], \\ \eta_A &:= \sum_{n \in \mathbb{N}_1} 1_A(X_n). \end{aligned} \tag{14}$$

Definition A.2 (Irreducibility)**.** *T is called* irreducible *if there exists a non-trivial σ-finite measure ϕ on $\mathcal{X}$ such that for $A \in \mathcal{B}_{\mathcal{X}}$ we have the implication:*

$$\phi(A) > 0 \quad \Longrightarrow \quad \forall x \in \mathcal{X}. \quad L(A|x) > 0. \tag{15}$$

The statement from Meyn & Tweedie (2012a) Prp. 4.2.2 allows for the following remark.

Remark A.3 (Maximal irreducibility measure)**.** *If T is irreducible then there always exists a non-trivial σ-finite measure ψ that is maximal (in the terms of absolute continuity) among all those ϕ with property 15. Such a ψ is unique up to equivalence (in terms of absolute continuity) and is called a* maximal irreducibility measure of T*. For such a ψ we introduce the notation:*

$$\mathcal{B}_{\mathcal{X}}^T := \{A \in \mathcal{B}_{\mathcal{X}} \mid \psi(A) > 0\}. \tag{16}$$

Note that $\mathcal{B}_{\mathcal{X}}^T$ does not depend on the choice of a maximal irreducibility measure ψ due to their equivalence. With this notation we then have for irreducible T:

$$A \in \mathcal{B}_{\mathcal{X}}^T \quad \Longrightarrow \quad \forall x \in \mathcal{X}. \quad L(A|x) > 0. \tag{17}$$

Definition A.4 (Harris recurrence)**.** *T is called* Harris recurrent *if T is irreducible and we have the implication:*

$$A \in \mathcal{B}_{\mathcal{X}}^T \quad \Longrightarrow \quad \forall x \in \mathcal{X}. \quad L(A|x) = 1. \tag{18}$$

Definition A.5 (Invariant probability measures)**.** *An* invariant probability measure (ipm) *of T is a probability measure μ on $\mathcal{X}$ such that:*

$$T \circ \mu = \mu. \tag{19}$$

On measurable sets this can equivalently be re-written as:

$$\forall A \in \mathcal{B}_{\mathcal{X}}. \qquad \int_{\mathcal{X}} T(A|x)\, \mu(dx) = \mu(A). \tag{20}$$

Remark A.6. *Note that a general Markov kernel T can have either no, exactly one or many invariant probability measures.*

For irreducible T we have the following results from Meyn & Tweedie (2012a) Prp. 10.1.1, Thm. 10.4.4, 18.2.2, concerning existence and uniqueness of invariant probability measures.

Theorem A.7 (Existence and uniqueness of invariant probability measures)**.** *Let T be irreducible.*

1. *Then T has at most one invariant probability measure μ; and:*
2. *the following are equivalent:*
   (a) *T has an invariant probability measure μ;*
   (b) *the following implication holds for $A \in \mathcal{B}_{\mathcal{X}}$:*

$$\begin{aligned} A \in \mathcal{B}_{\mathcal{X}}^T \quad &\Longrightarrow \quad \forall x \in \mathcal{X}. \\ &\limsup_{n \to \infty} T^n(A|x) > 0. \end{aligned} \tag{21}$$

We have the following properties of invariant probability measures for irreducible T. These are cited from Meyn & Tweedie (2012a) Thm. 9.1.5, Prp. 10.1.1, Thm. 10.4.4, 10.4.9, 10.4.10, and, Scheutzow & Schindler (2021) Prp. A.1, Lem. 3.2.

Theorem A.8 (Properties of irreducible Markov kernels with invariant probability measures)**.** *Let T be irreducible with invariant probability measure μ. Then the following statements hold:*

1. *μ is a maximal irreducibility measure for T.*
2. *μ satisfies the following condition for every $A \in \mathcal{B}_{\mathcal{X}}^{T}$ and $B \in \mathcal{B}_{\mathcal{X}}$:*
$$\mu(B) = \int_A \mathbb{E}_x \left[\sum_{n=1}^{\tau_A} 1[X_n \in B] \right] \mu(dx), \quad \tau_A := \inf \{ n \in \mathbb{N}_1 \mid X_n \in A \}. \tag{22}$$
3. *There exists a measurable set $\mathcal{H} \in \mathcal{B}_{\mathcal{X}}^{T}$ with $\mu(\mathcal{H}) = 1$ such that:*
$$\forall x \in \mathcal{H}. \quad T(\mathcal{H}|x) = 1, \tag{23}$$
T restricted to $\mathcal{H}$, $T : \mathcal{H} \dashrightarrow \mathcal{H}$, is well-defined and Harris recurrent (with invariant probability measure μ).

Definition A.9 (Aperiodicity)**.** *Let T be irreducible. Then T is called:*

1. periodic *if there exists $d \geq 2$ pairwise disjoint sets $A_1, \ldots, A_d \in \mathcal{B}_{\mathcal{X}}^{T}$, such that for every $j = 1, \ldots, d$, we have:*
$$\forall x \in A_j. \quad T(A_{j+1 (\mathrm{mod}\, d)}|x) = 1; \tag{24}$$
2. aperiodic *if T is not periodic.*

With these notation we have the following convergence theorems, see Meyn & Tweedie (2012a) Thm. 13.3.3, 17.0.1, and, Scheutzow & Schindler (2021) Thm. 2.16, 2.17, Assm. 2.12, Prp. 2.2.

Theorem A.10 (Strong Markov chain convergence theorem)**.** *Let μ be a probability measure on $\mathcal{X}$. Then the following are equivalent:*

1. *T is aperiodic and Harris recurrent and μ is an invariant probability measure for T.*
2. *For every $x \in \mathcal{X}$ we have the convergence in total variation norm:*
$$\lim_{n \to \infty} \mathrm{TV}(T_x^n, \mu) = 0. \tag{25}$$

Furthermore, if this is the case, then for every $g \in L^1(\mu)$ and every starting point $x \in \mathcal{X}$ we have the convergences:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k) = \mathbb{E}_\mu[g] \quad P_x\text{-a.s.} \tag{26}$$

Theorem A.11 (Markov chain convergence theorem)**.** *Let μ be a probability measure on $\mathcal{X}$. Then the following are equivalent:*

1. *T is aperiodic and irreducible and μ is an invariant probability measure for T.*
2. *For every $x \in \mathcal{X}$ we have:*
$$\lim_{n \to \infty} \mathrm{TV}(T_x^n, \mu) < 1, \tag{27}$$
and, for μ-almost-all $x \in \mathcal{X}$ we have the convergence in total variation norm:
$$\lim_{n \to \infty} \mathrm{TV}(T_x^n, \mu) = 0. \tag{28}$$

Furthermore, if this is the case, then for every $g \in L^1(\mu)$ and μ-almost-all starting points $x \in \mathcal{X}$ we have the convergences:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k) = \mathbb{E}_\mu[g] \quad P_x\text{-a.s.} \tag{29}$$

We now want to investigate under which conditions we can achieve irreduciblity, aperiodicity or Harris recurrence. We first cite the results of Asmussen & Glynn (2010) Thm. 1 and Cor. 1.

Theorem A.12 (Harris recurrence via irreducibility and density)**.** *Let T be irreducible with invariant probability measure μ. Further, assume that T has a density w.r.t. an irreducibility measure ϕ, i.e.:*

$$T(A|x) = \int_A t(y|x)\,\phi(dy), \tag{30}$$

with a jointly measurable $t : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$. Then ϕ is a maximal irreducibility measure for T, μ has a strictly positive density w.r.t. ϕ and T is Harrris recurrent.

Corollary A.13 (Harris recurrence via irreducibility and Metropolis-Hastings form)**.** *Let T be irreducible with invariant probability measure μ. Further, assume that T is of Metropolis-Hastings form w.r.t. an irreducibility measure ϕ:*

$$T(A|x) = (1 - a(x)) \cdot 1_A(x) + \int_A a(y|x) \cdot q(y|x)\,\phi(dy), \tag{31}$$

with jointly measurable $a, q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ and $a(x) > 0$ for every $x \in \mathcal{X}$. Note that: $a(x) = \int a(y|x) \cdot q(y|x)\,\phi(dy)$. Then ϕ is a maximal irreducibility measure for T, μ has a strictly positive density w.r.t. ϕ and T is Harrris recurrent.

We now have all ingredients to derive the following criteria for the strong Markov chain convergence theorem A.10 to apply:

Corollary A.14 (Criterion for convergence via positive density)**.** *Let ϕ be a non-trivial σ-finite measure on $\mathcal{X}$ such that T has a strictly positive jointly measurable density $t : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{>0}$ w.r.t. ϕ:*

$$T(A|x) = \int_A t(y|x)\,\phi(dy), \tag{32}$$

then T is irreducible, aperiodic and ϕ is a maximal irreducibility measure for T.

If, furthermore, T has an invariant probability measure μ then μ has a strictly positive density w.r.t. ϕ, T is Harris recurrent and the strong Markov chain convergence theorem A.10 applies.

Corollary A.15 (Criterion for convergence via positive Metropolis-Hastings form)**.** *Let μ be an invariant probability measure of T. Further, assume that T is of Metropolis-Hastings form w.r.t. a non-trivial σ-finite measure ϕ:*

$$T(A|x) = (1 - a(x)) \cdot 1_A(x) + \int_A a(y|x) \cdot q(y|x)\,\phi(dy), \tag{33}$$

with strictly positive jointly measurable $a, q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{>0}$ such that for every $x \in \mathcal{X}$ we have that:

$$a(x) := \int a(y|x) \cdot q(y|x)\,\phi(dy) \overset{!}{\in} (0, 1). \tag{34}$$

Then ϕ is a maximal irreducibility measure for T, μ has a strictly positive density w.r.t. ϕ, T is aperiodic, Harrris recurrent and the strong Markov chain convergence theorem A.10 applies.

Corollary A.16 (Criterion for convergence on countable spaces)**.** *Let $\mathcal{X}$ be a countable space, i.e. finite or countably infinite. Let T be irreducible with invariant probability measure μ such that for all $x \in \mathcal{X}$ with $\mu(\{x\}) > 0$ we also have $T(\{x\}\,|x) > 0$. Then T is aperiodic and Harris recurrent and the strong Markov chain convergence theorem A.10 applies.*

B Additional Clarifications and Discussion

In this section, we provide additional clarifications and discussion.

B.1 Definition of Long-term Fairness

We provide an overview of how prior research's fairness formulations relate to our definitions of long-term fair targets.

First, our framework aims to attain a state of long-term fairness. This entails that fairness formulations should be met in the long term and, importantly, once achieved, maintained consistently. Our goal differs fundamentally from approaches that aim to fulfill fairness at each time step. In this regard, (D'Amour et al., 2020) compare agents optimizing for short-term goals - e.g., a profit-maximization agent to an equality of opportunity fair agent and measure the long-term (in)equality of the initial credit score distribution across groups - without imposing it on the agents.

Prior work on long-term fairness introduces parity of return (Chi et al., 2022), which requires equal discounted rewards accumulated by the decision-maker over time, where the reward could be defined as the ratio between true positive and overall positive decisions. Wen et al. (2021) define long-term demographic parity (equal opportunity) as asking the cumulative expected individual rewards to be on average equal for (qualified members of) demographic groups. (Yin et al., 2023) aim to maximize the accumulated reward subject to accumulated unfairness (utility) constraint in a finite time horizon. The reward combines true positive and true negative rates, while the authors consider different (un)fairness measures: demographic parity, equal opportunity, and equal qualification rate. (Yu et al., 2022) formulate a (short-term) fairness metric (e.g., equality of opportunity) as a function of the state and increase its enforcement over time.

Our framework provides the capability to enforce these fairness and reward considerations, specifically we allow for feature complex objective functions (see § 6.1) as well as imposing feature (qualification) equality § 6.2) and group fairness criteria in the long-term (see § 6.3) for infinite time-horizons. Note that the formulation of a fair state is not limited to the possible fairness objectives and constraints discussed in § 6. Rather, we exemplify in that section that our framework can capture fairness objectives well-established in prior work (in addition to the above cited: Zhang et al. (2020); Liu et al. (2018); Dwork et al. (2012); Hardt et al. (2016b)).

B.2 Assumption of Known or Estimatable Dynamics

Our work takes a structured approach by separating the estimation problem (of the Markov kernel i.e., the dynamics) from the policy learning process. We recognize that the estimation problem itself is a significant challenge and requires careful attention and, as commented in § 8, is the subject of a different line of active research and thus outside the scope of this paper.

The quality of the dynamics estimation heavily relies on the quality and quantity of the available temporal data, the complexity of the environment, and the estimation methods (as it does, e.g., for model-based reinforcement learning). Estimation of dynamics / Markov kernels is an active research field Sherlaw-Johnson et al. (1995); Craig & Sendi (2002); Wu et al. (2018); Sun et al. (2019) and our method can benefit from the advances made in the field. If temporal data is available, estimating dynamics may even prove to be faster and more data-efficient than learning them through interactions. We exemplify estimating dynamics in additional results in Appendix E.6.

Further, within our framework and application, dynamics we describe consequences of decisions on individuals' features. The dynamics in the lending example of our experiments are determined by the credit score maker's policy on how scores are updated in response to (un)paid credits. Though our framework is not limited to this, dynamics - themselves depending on a statistical/rule-based/ML model - may be accessible or much simpler to estimate than complex human behavior.

B.3 Existence of a Fair Stationary Distribution

Our approach also serves to determine whether a stationary distribution exists. In situations where a fair policy does indeed exist, our optimization problem (OP) is designed to effectively discover it. If a solution to our optimization problem does not exist, it implies that alternative methods (including, e.g., reinforcement learning), would also not find a policy inducing and maintaining the targeted fair stationary distribution under the same modeling assumptions. This stems from the fact that if the current state is fair, any alternative approach would still need to address the stationary equation (3) to maintain that state. This discovery can offer valuable insights to practitioners, prompting them to

explore different perspectives on long-term fairness. For instance, this might involve revising non-stationary long-term fairness objectives, such as addressing oscillating long-term behaviors (Zhang et al., 2020). Alternatively, practitioners could consider redefining the targeted fair state that allows for stationary. By shedding light on these possibilities, our approach contributes to a deeper understanding of the dynamics and long-term fairness considerations.

B.4 Choice of Dataset

Our current experiment focuses on a single simulation setup, specifically centered around loan repayment. At the same time, we provide results for varying dynamics and initial distributions, essentially simulating different datasets of the same generative model. Note also that we provide an example of how the framework can be applied to a different generative model in Appendix F. Finally, focusing on a single generative model (Zhang et al., 2020) and a single guiding example is in line with prior published work (Zhang et al., 2020; Liu et al., 2018; Creager et al., 2020; Wen et al., 2021) with the loan example used widely by previous work on long-term fairness (D'Amour et al., 2020; Liu et al., 2018; Creager et al., 2020; Wen et al., 2021; Yu et al., 2022).

B.5 Opportunities and Limitations of Time-invariant Policies

Our framework yields a single fixed, i.e., time-invariant policy. When the dynamics are constant, and policy learning and estimation of the dynamics occur simultaneously (as in reinforcement learning), then the learned policy requires frequent updates as more data becomes available. Our paper takes a different approach by separating the estimation problem (of the Markov kernel i.e., the dynamics) from the policy learning process and therefore does not require updating the policy. We believe that this holds several advantages, particularly in terms of predictability and trustworthiness. A fixed policy provides a consistent decision-making framework that stakeholders can anticipate and understand contributing to trustworthiness. In addition, a fixed policy simplifies operational processes, such as implementation and maintenance efforts, potentially leading to more efficient and effective outcomes.

When the dynamics vary with time, we can no longer rely on a single time-invariant policy for an infinite time horizon. If, however, the changes are slow and the dynamics remain constant within certain time intervals, our approach remains effective within the time intervals. Whenever the dynamics change, our approach would require re-estimating the dynamics and solving the optimization problem again to obtain a new policy. In this way, our method adapts to changing conditions and maintains its effectiveness over time. However, when dynamics change rapidly, the adaptability of any method is limited.

B.6 Modeling Choice

Our intention in developing a framework for long-term fair policy learning is to provide a versatile approach that could be applied across various contexts. While models serve as simplified representations of complex systems, they allow us to analyze phenomena otherwise incomprehensible. Our choice of utilizing Markov Chains as a modeling tool is a reflection of this principle. Markov Chains are chosen for their wide application in understanding dynamic processes. For example, the field of Reinforcement Learning relies on Markov Decision Processes (MDPs), a specific kind of Markov Chain. The proposed modeling framework can indeed be adapted to a variety of different scenarios and we provide an example of a different scenario / generative model in Appendix F.

C On Long-term Targets

In this section, we provide additional details regarding the targeted fair states introduced in § 6.

C.1 On Minimax Objectives

In § 6.2, it was mentioned that egalitarian distributions may not always be efficient, and there are cases where minimizing the maximum societal risk is more desirable to prevent unnecessary harm. We elaborate on this concept in the following. While egalitarian allocations can align with societal

values, they are generally considered Pareto inefficient (Pazner, 1975). In certain scenarios, policy-makers may be interested in minimizing the maximum risk within a society (Barsotti & Koçer, 2022). This approach aims to prevent unnecessary harm by reducing the risk for one group without increasing the risk for another (Martinez et al., 2020). For instance, in the context of hiring, instead of equalizing the group-dependent repayment rates $\mathcal{Q}(\pi, s)$, a policy-maker may be interested in minimizing the maximum default risk $1 - \mathcal{Q}(\pi, s)$ across groups. In other words, their objective could be $J_{\text{LT}} := \min_s -(1 - \mathcal{Q}(\pi, s))$, rather than aiming for equal default or repayment rates.

C.2 Policy constraints

In § 6.3, we mentioned that it is possible to incorporate constraints on the type of policy being searched for. These constraints could be put on the policy independent of the stationary distribution. We provide an example here. If the features exhibit a monotonic relationship, where higher values of X_t tend to result in a higher probability of a positive outcome of interest $\ell(Y = 1 \mid x, s)$, we may also be interested in a monotonous policy. A monotonous policy assigns higher decision probabilities as X_t increases. In such cases, we can impose the additional constraint $\pi(k, s) \geq \pi(x, s), \forall k \geq x, s$.

D Simulation Details

In this section, we present the details of the experiments and simulations in § 7.

D.1 Solving the Optimization Problem

Our framework can be thought of as a three-step process. First, just as previous work on algorithmic fairness empowers users to choose fairness criteria, our framework allows users to define the characteristics of a fair distribution applicable in their decision-making context (see § 6). The second step involves transforming the definition of fair characteristics into an optimization problem (OP). The third step consists of solving the OP. Given the nature of our optimization problem, which is linear and constraint-based, we can employ any efficient black-box optimization methods for this class of problems. Note that the OP seeks to find a policy π that induces a stationary distribution μ, which adheres to the previously defined fairness targets. As detailed in § 7, in the search of π, we first compute group-dependent kernel T^s_π, which is a linear combination of assumed/estimated dynamics and distributions and π. We then compute the group-dependent stationary distribution μ^s_π via eigendecomposition.

Solving the Optimization Problem for Finite State Spaces In our guiding example and the corresponding simulation, we consider a time-homogeneous Markov chain $(\mathcal{Z}, P)$ with a finite state space $\mathcal{Z}$ (e.g., credit score categories). Consequently, the convergence constraints C_{conv} are determined by the *irreducibility* and *aperiodicity* properties of the corresponding Markov kernel (see § 4).

Recall from Def. 4.2 that a time-homogeneous Markov chain is considered *irreducible* if, for any two states $z, w \in \mathcal{Z}$, there exists a $t > 0$ such that $P^t(z, w) > 0$, where $P^t(z, w) = \mathbb{P}(Z_t = w \mid Z_0 = z)$ represents the probability of going from z to w in t steps.

To ensure irreducibility in our optimization problem, we impose the condition $\sum_{i=1}^n P^n > \mathbf{0}$, where $n = |\mathcal{Z}|$ is the number of states and $\mathbf{0}$ denotes the matrix with all entries equal to zero. We can demonstrate that this implies irreducibility through a proof by contradiction: Suppose that $\sum_{i=1}^n P^n > \mathbf{0}$, but for all $t \in \{1, 2, \ldots, n\}$, we have $P^t(z, w) = 0$ for all z and w. Then $\sum_{t=1}^n P^n = 0$, which contradicts the initial condition. Consequently, if $\sum_{i=1}^n P^n > \mathbf{0}$, it follows that there exists a $t > 0$ such that $P^t(z, w) > 0$.

To satisfy *aperiodicity* in our optimization we require that the diagonal elements of the transition matrix are greater than zero: $P(z, z) > 0$ for all z, where $P(z, z)$ represents the diagonal elements of the Markov kernel P. Recall from Def. 4.3, that we denote $R(z) = \{t \geq 1 : P^t(z, z) > 0\}$ to be the set of return times from $z \in \mathcal{Z}$, where $P^t(z, z)$ represents the probability of returning to state x after t steps. The Markov chain is aperiodic if and only if the greatest common divisor (gcd) of $R(z)$ is equal to 1: $gcd(R(z)) = 1$ for all z in $\mathcal{Z}$. If $P^1(z, z) > 0$ for all z, then $t = 1$ is in $R(z)$, which means that the gcd of $R(z)$ is equal to 1.

Following Theorem 4.1, a sufficient condition for convergence to the unique stationary distribution is the positivity of the transition matrix P, where all elements are greater than zero. Therefore, if we assume the transition matrix to be positive, we do not need to impose the *irreducibility* and *aperiodicity* constraints mentioned above. In our experiments, for the sake of simplicity, we ensure that the transition matrix P is positive, meaning that all its elements are greater than zero. Specifically, in our guiding example, this assumption implies that we assume $g(k \mid x, d, y, s) > 0$ for all d, s, y, x, k, while FICO data already yields $\ell(y \mid x, s) > 0$ for all y, x, s.

We compute the *stationary distribution* μ using eigendecomposition. Recall from Definition 3.2 that a stationary distribution of a time-homogeneous Markov chain $(\mathcal{Z}, P)$ is a probability distribution μ such that $\mu = \mu P$. More explicitly, for every $w \in \mathcal{Z}$, the following needs to hold: $\mu(w) = \sum_z \mu(z) \cdot P(z, w)$. If the transition matrix P is positive, $\mu = \mu P$ implies that μ is the eigenvector of P corresponding to eigenvalue 1. We then solve for the stationary distribution μ using linear algebra.

SLSQP Algorithm We solve optimization problems (5) and (6) using the Sequential Least Squares Programming (SLSQP) method Kraft (1988). SLSQP is a method used to minimize a scalar function of multiple variables while accommodating bounds, equality and inequality constraints and can be used for solving both linear and non-linear constraints. The algorithm iteratively refines the solution by approximating the objective function and constraints using quadratic model.

Specifically, SLSQP is designed to minimize scalar functions of one or more variables. In our case we are maximizing utility ($\pi^\star_{\texttt{EOP}}$) or qualifications ($\pi^\star_{\texttt{QUAL}}$) and searching for $\mathbb{P}(D = 1 \mid X = x, S = s)$ for all x and s, which are with $|X| = 4$ and $|S| = 2$, a total of 8 variables. Further, SLSQP can handle optimization problems with variable bounds. In our case, we set a minimum bound of 0 and a maximum bound of 1 as we are seeking for probabilities $\mathbb{P}(D = 1 \mid X = x, S = s)$ for all x and s. SLSQP can also handle both linear and non-linear equality and inequality constraints. In our example, where the state space is finite (i.e., X is categorical), all constraints are linear inequality or equality constraints. Finally, SLSQP uses a sequential approach, which means it iteratively improves the solution by solving a sequence of subproblems. This approach often converges efficiently, even for non-convex and non-linear optimization problems.

We use the SLSQP solver from scikit-learn[1] (Pedregosa et al., 2011) with step size eps $\approx 1.49 \times 10^{-10}$ and a max. number of iterations 200 and initialize the solver (warm start) with a uniform policy where all decisions are random, i.e., $\pi(D = 1 \mid x, s) = 0.5 \, \forall x, s$.

D.2 Assumed Dynamics

We now provide details about the assumed dynamics. Note that in our guiding example, we assume binary $s, y, d \in \{0, 1\}$ and four credit categories, i.e., we have $n = |\mathcal{X}| = 4$ states. For simplicity we assume the following notation: $T_{sdy} := g(k \mid x, d, y, s)$. T_{sdy} is a $n \times n$ transition matrix that describes the Markov chain, where the rows and columns are indexed by the states, and $T_{sdy}(x, k)$, i.e., the number in the x-th row and k-th column, gives the probability of going to state $X_{t+1} = k$ at time $t + 1$, given that it is at state $X_t = x$ at time t and given that $S = s$, $D_t = d$, $Y_t = y$.

One-sided Dynamics. For all one-sided dynamics (in § 7 and E.5) we assume:

[1] https://docs.scipy.org/doc/scipy/reference/optimize.minimize-slsqp.html

$$
\begin{aligned}
&T_{000} = T_{001} = T_{100} = T_{101} \\
&= \begin{bmatrix} 0.9 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.9 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.9 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.9 \end{bmatrix} \\
&T_{110} = T_{010} \\
&= \begin{bmatrix} 0.9 & 0.9 & 0.9 & 0.9 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \end{bmatrix}
\end{aligned} \tag{35}
$$

One-sided General. For the one-sided dynamics in § 7.1 we additionally assume dynamics T_{sdy} that depend on the sensitive attribute in addition to (35):

$$
T_{111} = \begin{bmatrix} 0.53333 & 0.03333 & 0.03333 & 0.03333 \\ 0.4 & 0.53333 & 0.03333 & 0.03333 \\ 0.03333 & 0.4 & 0.53333 & 0.03333 \\ 0.03333 & 0.03333 & 0.4 & 0.9 \end{bmatrix}
$$

$$
T_{011} = \begin{bmatrix} 0.33333 & 0.03333 & 0.03333 & 0.03333 \\ 0.6 & 0.33333 & 0.03333 & 0.03333 \\ 0.03333 & 0.6 & 0.33333 & 0.03333 \\ 0.03333 & 0.03333 & 0.6 & 0.9 \end{bmatrix}
$$

One-sided Slow. For the one-sided `slow` dynamics with results presented in E.5, we assume the following group-independent dynamics T_{sdy} in addition to (35):

$$
T_{011} = T_{111} = \begin{bmatrix} 0.53333 & 0.03333 & 0.03333 & 0.03333 \\ 0.4 & 0.53333 & 0.03333 & 0.03333 \\ 0.03333 & 0.4 & 0.53333 & 0.03333 \\ 0.03333 & 0.03333 & 0.4 & 0.9 \end{bmatrix}
$$

One-sided Medium. For the one-sided `medium` dynamics with results presented in E.5, we assume the following group-independent dynamics T_{sdy} in addition to (35):

$$
T_{011} = T_{111} = \begin{bmatrix} 0.33333 & 0.03333 & 0.03333 & 0.03333 \\ 0.6 & 0.33333 & 0.03333 & 0.03333 \\ 0.03333 & 0.6 & 0.33333 & 0.03333 \\ 0.03333 & 0.03333 & 0.6 & 0.9 \end{bmatrix}
$$

One-sided Fast. For the one-sided `fast` dynamics with results presented in E.5, we assume the following group-independent dynamics T_{sdy} in addition to (35):

$$
\begin{aligned}
&T_{011} = T_{111} \\
&= \begin{bmatrix} 0.13333 & 0.03333 & 0.03333 & 0.03333 \\ 0.8 & 0.13333 & 0.03333 & 0.03333 \\ 0.03333 & 0.8 & 0.13333 & 0.03333 \\ 0.033335 & 0.03333 & 0.8 & 0.9 \end{bmatrix}
\end{aligned}
$$

Two-sided Recourse Dynamics. For `recourse` dynamics we assume the following dynamics T_{sdy}. Specifically, we assume that dynamics are the same for both sensitive groups.

$$T_{000} = T_{001} = \begin{bmatrix} 0.7 & 0.03333 & 0.03333 & 0.03333 \\ 0.23333 & 0.7 & 0.03333 & 0.03333 \\ 0.03333 & 0.23333 & 0.7 & 0.03333 \\ 0.03333 & 0.03333 & 0.23333 & 0.9 \end{bmatrix}$$

$$T_{100} = T_{101} = \begin{bmatrix} 0.5 & 0.03333 & 0.03333 & 0.03333 \\ 0.43333 & 0.5 & 0.03333 & 0.03333 \\ 0.03333 & 0.43333 & 0.5 & 0.03333 \\ 0.03333 & 0.03333 & 0.43333 & 0.9 \end{bmatrix}$$

$$T_{010} = T_{011} = \begin{bmatrix} 0.9 & 0.9 & 0.9 & 0.9 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \end{bmatrix}$$

$$T_{110} = T_{111} = \begin{bmatrix} 0.33333 & 0.03333 & 0.03333 & 0.03333 \\ 0.6 & 0.33333 & 0.03333 & 0.03333 \\ 0.03333 & 0.6 & 0.33333 & 0.03333 \\ 0.03333 & 0.03333 & 0.6 & 0.9 \end{bmatrix}$$

Two-sided Discouraged Dynamics. For `discouraged` dynamics we assume the following dynamics T_{sdy}. Specifically, we assume that dynamics are the same for both sensitive groups.

$$T_{000} = T_{001} = \begin{bmatrix} 0.9 & 0.63333 & 0.13333 & 0.03333 \\ 0.03333 & 0.3 & 0.53333 & 0.23333 \\ 0.03333 & 0.03333 & 0.3 & 0.43333 \\ 0.03333 & 0.03333 & 0.03333 & 0.3 \end{bmatrix}$$

$$T_{100} = T_{101} = \begin{bmatrix} 0.9 & 0.43333 & 0.13333 & 0.03333 \\ 0.03333 & 0.5 & 0.33333 & 0.23333 \\ 0.03333 & 0.03333 & 0.5 & 0.23333 \\ 0.03333 & 0.03333 & 0.03333 & 0.5 \end{bmatrix}$$

$$T_{010} = T_{011} = \begin{bmatrix} 0.9 & 0.9 & 0.9 & 0.9 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \\ 0.03333 & 0.03333 & 0.03333 & 0.03333 \end{bmatrix}$$

$$T_{110} = T_{111} = \begin{bmatrix} 0.33333 & 0.03333 & 0.03333 & 0.03333 \\ 0.6 & 0.33333 & 0.03333 & 0.03333 \\ 0.03333 & 0.6 & 0.33333 & 0.03333 \\ 0.03333 & 0.03333 & 0.6 & 0.9 \end{bmatrix}$$

D.3 Setup of different runs

Different Random Initial Distributions. In order to generate the results presented in § 7.1, we solve the optimization problem (5) once, using $\epsilon = 0.01$ and $c = 0.08$, and obtain the optimal policy $\pi^{\star}_{\texttt{EOP}}$. Subsequently, we perform simulations for 10 different random initial distributions $\mu_0(x \mid s)$, where we observe the behavior of $\pi^{\star}_{\texttt{EOP}}$ for the assumed dynamics `one-sided` over a duration of 200 steps. In line with the Markov convergence theorem, all of these simulations yield the same stationary distribution.

Different Dynamic Types. To obtain results in § 7.1, we address two optimization problems: (5) and (6). For each set of dynamics, we solve these problems independently, resulting in different values for $\pi^{\star}_{\texttt{EOP}}$ and $\pi^{\star}_{\texttt{QUAL}}$ respectively. Subsequently, we utilize the FICO distribution as the initial distribution $\mu_0(x \mid s)$ and simulate the feature distribution for each policy over 200 steps, assuming the specified dynamics.

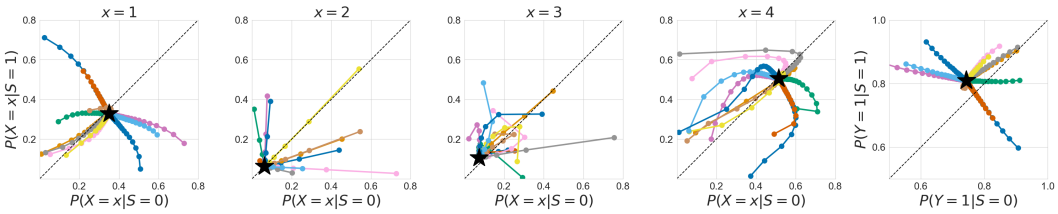


Figure 4: Convergence of feature distributions for $\pi^{\star}_{\texttt{EOP}}$ for different random starting distributions (colors) to unique stationary distributions $\mu = \star$. Trajectories over 200 time steps. $c = 0.8$, $\epsilon = 0.01$.

D.4 Computational Resources and Run Time

Computational Resources All experiments were conducted on a MacBook Pro (Apple M1 Max chip). Since we can efficiently solve the optimization problem, these experiments are executed on standard hardware, eliminating the necessity for using GPUs.

Run Time The optimization problems to find long-term policies in all experiments within this paper were consistently solved in under 10 seconds. Regarding the training of short-term fair policies on 5000 samples, the run times were approximately 20-23 minutes: 1245.92 seconds for `short-EOP` ($\lambda = 1$), 1244.25 seconds for `short-EOP` ($\lambda = 2$), and 1380.50 seconds for `short-MAXUTIL`.

E Additional Results

In this section, we provide additional results related to the results discussed in § 7. Our analysis centers around our guiding example, employing the data distributions sourced from FICO (Reserve, U. F., 2007) unless otherwise specified. The structure of this section is as follows:

- In § E.1 we provide additional results for different starting distributions.
- In § E.2 we provide additional results for the comparison to short-term policies.
- In § E.3 we provide additional results for varying the fairness threshold ϵ for our policy.
- In § E.4 we provide additional results for the different dynamic types (`one-sided`, `recourse`, `discouraged`) that we introduced in the main paper.
- In § E.5 we provide additional results for varying the speed at which feature changes occur (`slow`, `medium`, `fast`).
- In § E.6 we provide additional results for first sampling from FICO data and then estimating the distributions under partially observed labels.

E.1 Different Initial Starting Distributions

We provide additional results for the results shown § 7.1, where we run simulations on 10 randomly sampled initial feature distributions $\mu_0(x \mid s)$, setting $\epsilon = 0.01, c = 0.8$. In addition to the results shown in the main paper, we here display in Figure 4 the resulting trajectories of all feature distributions.

E.2 Comparison to Static Policies

We provide additional results comparing our long-term policy to short-term policies.

Static Policy Training. The short-term policies are logistic regression models implemented using PyTorch. The forward method computes the logistic sigmoid of a linear combination of the input features, while the prediction method applies a threshold of 0.5 to the output probability to make binary predictions. The training process is carried out via gradient descent, with the train function

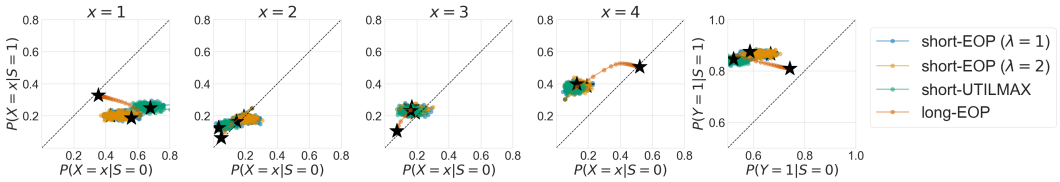


Figure 5: Convergence of feature distributions for our long-term `long-EOP` ($\pi^{\star}_{\texttt{EOP}}$) and the static policies (unfair: `short-MAXUTIL`, fair: `short-EOP` ($\lambda = 2$). Trajectories over 200 time steps. $c = 0.8$, $\epsilon = 0.026$. Last distribution values are marked with $\star$.

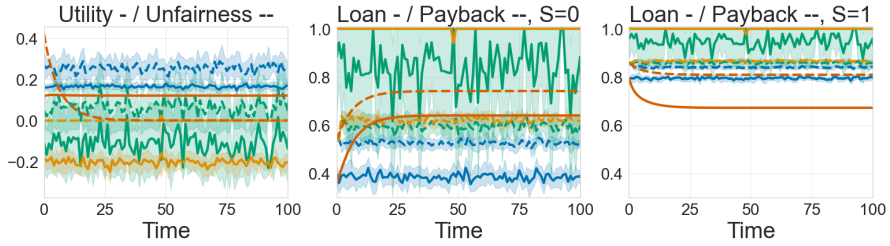


Figure 6: Results for our long-term `long-EOP` ($\pi^{\star}_{\texttt{EOP}}$) and the static policies (unfair: `short-MAXUTIL`, fair: `short-EOP` ($\lambda = 2$). . Top Left: Utility (solid, $\uparrow$) with $c = 0.8$ and EOP-Unfairness (dashed, $\downarrow$). Top right / Bottom left: Loan (solid) and payback probability (dashed) per policy and sensitive S.

optimizing a specified loss function. The `short-MAXUTIL` policy is trained using a binary cross-entropy loss. The fairness is enforced using a Lagrangian approach ($\lambda = 2$). The `short-EOP` policy is trained using a binary cross-entropy loss and regularization terms measuring equal opportunity unfairness with λ as hyperparameters controlling the trade-off between predictive accuracy and fairness. Training is performed for 2000 epochs with a learning rate of 0.05. We display results over 10 random initializations. The experiments in the main paper are shown for `short-EOP` with $\lambda = 2$. We show in the following results for different λ.

Feature and Outcome Trajectories. Figure 5 presents the trajectories of our long-term `long-EOP` ($\pi^{\star}_{\texttt{EOP}}$) and the static policies (unfair: `short-MAXUTIL`, fair: `short-EOP` ($\lambda = 2$)) over 200 time steps for a single short-term policy seed. We observe that our long-term policy converges to a stationary distribution and remains there once it has found it. In contrast, the trajectories of the short-term policies display non-stationarity, covering a wide range of distributions, as evidenced by the overlapping region. This indicates that the short-term policies exhibit a high variance and do not stabilize into a stationary distribution.

Utility, Fairness and Loan and Repayment Probabilities. Figure 6 (top left) displays $\mathcal{U}$ and `EOPUnf` over the first 100 time steps. We observe that short-term policies, which are updated at each time step, tend to exhibit greater variance compared to the long-term policy, which remains fixed at $t = 0$ - even as the underlying data distribution evolves in response to decision-making. Among the two short-term fair policies, the fairer one ($\lambda = 2$) approaches nearly zero unfairness, whereas the less fair one ($\lambda = 1$) displays a higher level of unfairness. Specifically, the more fair policy ($\lambda = 2$) reaches a low (negative) utility, while the less fair one ($\lambda = 1$) maintains a higher (though still negative) utility. The unfair short-term policy (`UTILMAX`) achieves positive utility but does so at the cost of a high level of unfairness. This highlights the trade-off between fairness and utility that short-term policies encounter. Conversely, our long-term fair policy maintains a level of unfairness close to zero while experiencing only a modest reduction in utility compared to the unfair short-term policy. This underscores our policy's capacity to attain long-term fairness while ensuring a higher level of utility, leveraging the long-term perspective to effectively shape the population distribution.

Figure 6 (top right, bottom left) presents the loan probability $\mathbb{P}(D = 1 \mid S = s)$ and payback probability $\mathbb{P}(Y = 1 \mid S = s)$ for non-privileged ($S = 0$) and privileged ($S = 1$) groups. In

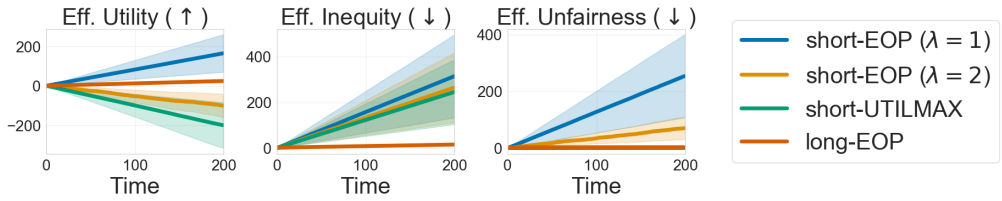


Figure 7: Results for our long-term `long-EOP` ($\pi^{\star}_{\text{EOP}}$) and the static policies (unfair: `short-MAXUTIL`, fair: `short-EOP` ($\lambda = 2$). Effective (cumulative) utility $\mathcal{U}$, inequity $\mathcal{I}$, and (EOP) unfairness `EOPUnf` for different policies.

addition to the results presented in the main paper (Figure 2b), we observe a difference between the two short-term fair policies in our analysis in this appendix. The more equitable policy ($\lambda = 2$) achieves a low level of unfairness by granting loans with a probability of 1 to individuals across all social groups. The less equitable policy ($\lambda = 1$) provides loans to the underprivileged group with an average probability of approximately 0.85, while the privileged group receives loans at an average probability of around 0.9.

Crucially, the less equitable policy ($\lambda = 1$) exhibits a much higher variability in loan approval probabilities for the underprivileged group across different time steps compared to the privileged group. This highlights that unfairness does not solely manifest at the mean level but also in the variability across time. Both policies tend to grant loans at probabilities exceeding the actual repayment probabilities within the population. This suggests an "over-serving" phenomenon, implying that the policies on average extend loans to individuals who may not meet the necessary qualifications for borrowing.

In contrast, our policy maintains stability and converges to a low difference in loan approval probabilities between groups without significant temporal variance. Importantly, our loan approval probabilities remain below the loan repayment (as for the short-term unfair policy (`UTILMAX`)) probabilities, indicating that, on average, the policies are extending loans to individuals who are indeed eligible for them. In addition, for our policy, the gap between loan provision and repayment probabilities is similar across sensitive groups.

Effective Utility, Inequity and Unfairness. Figure 7 illustrates effective (accumulated) measures of utility, inequity, and (EOP) unfairness over time for the different policies, where results for static policies are reported over 10 random initializations. We observe that the short-term unfair policy (`short-UITLMAX` consistently accumulates the highest utility across all dynamics, while simultaneously maintaining a high level of effective unfairness and inequality. Conversely, the short-term fair policies (`short-EOP`($\lambda = 1$) and ($\lambda = 2$)) exhibit negative effective utility, but they do achieve lower levels of effective fairness and inequity.

For our long-term policy (`long-EOP`), we find that it accumulates positive utility over time. Although its utility remains below that of the short-term unfair policy, our policy exhibits very low levels of effective unfairness. Importantly, it also yields minimal accumulated inequity, even though it was not specifically optimized for this.

Analyzing the cumulative effects of policies is essential for evaluating the long-term impact of each policy choice. This analysis can, for instance, help determine whether investing in fairness pays off in the long-term and whether sacrificing short-term fairness in the initial stages ultimately benefits society in the long run.

E.3 Different Fairness Levels

We provide additional results, where we use the initial distribution $\mu_0(x \mid s)$ from FICO and solve the optimization problem (5) for four different fairness levels ϵ. This results in four policies $\pi^{\star}_{\text{EOP}}$.

Feature and Outcome Trajectories. Figure 8 presents the trajectories of $\pi^{\star}_{\text{EOP}}$ over 200 time steps for different fairness thresholds ϵ. We observe that although the convergence process, time, and final stationary distribution ($\star$) are very similar for different targeted fairness levels.

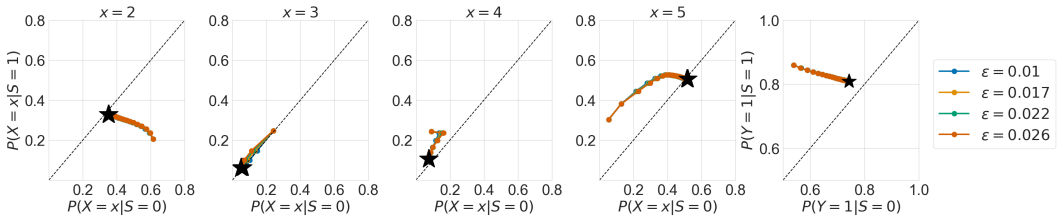


Figure 8: Convergence of feature distributions for $\pi^{\star}_{\texttt{EOP}}$ for different fairness thresholds ϵ to unique stationary distributions $\mu = \star$. Trajectories over 200 time steps. $c = 0.8$.

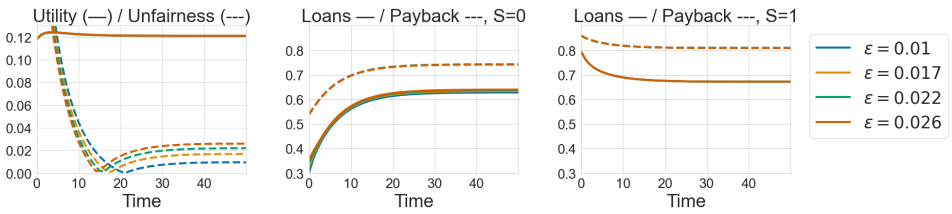


Figure 9: Results for different ϵ-EOP-fair $\pi^{\star}_{\texttt{EOP}}$. Top Left: Utility (solid, $\uparrow$) with $c = 0.8$ and EOP-Unfairness (dashed, $\downarrow$). Top right / Bottom left: Loan (solid) and payback probability (dashed) per policy and sensitive S.

Utility and Loan and Repayment Probabilities. Figure 9 (top left) displays $\mathcal{U}$ and `EOPUnf` over the first 50 time steps (until convergence). We observe that all policies converge to a similar utility level while maintaining their respective ϵ level, confirming the effectiveness of our optimization problem. Figure 9 (top right, bottom left) presents the loan probability $\mathbb{P}(D = 1 \mid S = s)$ and payback probability $\mathbb{P}(Y = 1 \mid S = s)$ for non-privileged ($S = 0$) and privileged ($S = 1$) groups. While the probabilities across sensitive groups ultimately stabilize close together in the long term, the initial 20 steps exhibit a large difference in loan and payback probabilities. Optimizing for long-term goals may thus lead to unfairness in the short term, and it is important to carefully evaluate the potential impact of this on public trust in the policy.

E.4 Different Dynamic Types

Results in this subsection are for different dynamic types: `one-sided`, `recourse`, and `discouraged`. See D.2 for more details on these specific dynamics. We solve both optimization problems for each of the three dynamics, where solving (5) provides $\pi^{\star}_{\texttt{EOP}}$ and solving (6) provides $\pi^{\star}_{\texttt{QUAL}}$.

Feature and Outcome Trajectories. Figure 10 presents the trajectories of $\pi^{\star}_{\texttt{EOP}}$ and $\pi^{\star}_{\texttt{QUAL}}$ over 200 time steps for different types of dynamics. We observe that although the initial distribution remains unchanged, the convergence process, time, and final stationary distribution ($\star$) differ depending on the dynamics. Notably, the stationary distribution of $\pi^{\star}_{\texttt{QUAL}}$ appears to be similar for `one-sided` and `discouraged` dynamics. On the other hand, the results for all other dynamics and policies demonstrate distinct but relatively close outcomes.

Utility, Fairness and Loan and Repayment Probabilities. Figure 11 showcases the group-dependent probabilities of receiving a loan, $\mathbb{P}(D_t = 1 \mid S = s)$, and repayment, $\mathbb{P}(Y_t = 1 \mid S = s)$, for both the non-privileged ($S = 0$) and privileged ($S = 1$) groups. The probabilities are displayed for the convergence phase (first 50 time steps) for policies $\pi^{\star}_{\texttt{EOP}}$ and $\pi^{\star}_{\texttt{QUAL}}$ across dynamics types. When the payback probabilities are higher compared to the loan probabilities, it suggests an underserved community where fewer credits are granted than would be repaid. In the case of `one-sided` dynamics, we find that for $\pi^{\star}_{\texttt{EOP}}$, the loan and repayment probabilities are relatively close to each other at each time step. However, for $\pi^{\star}_{\texttt{QUAL}}$, the gap between repayment and loan probabilities widens as time progresses. At convergence, both sensitive groups exhibit a repayment rate of approximately 0.8, while the loan-granting probability is around 0.4. This suggests that, in

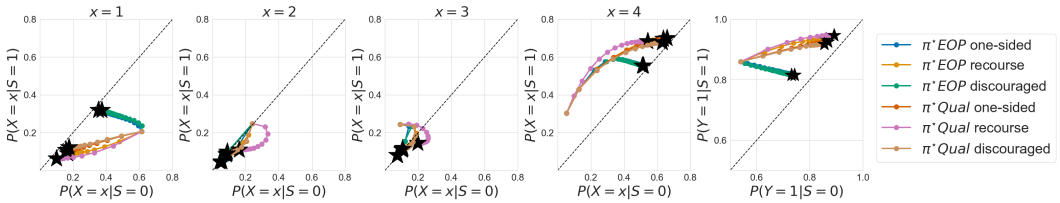


Figure 10: Convergence of $\pi^\star_{\text{EOP}}$ and $\pi^\star_{\text{QUAL}}$ for different type of dynamics towards different unique stationary distributions $\mu = \star$. Trajectories over 200 time steps. Top four plots: feature distribution μ_t. Bottom left: distribution of the outcome of interest. Equal feature/outcome distribution dashed. Initial distribution μ_0 =FICO, $c = 0.8$, $\epsilon = 0.01$.

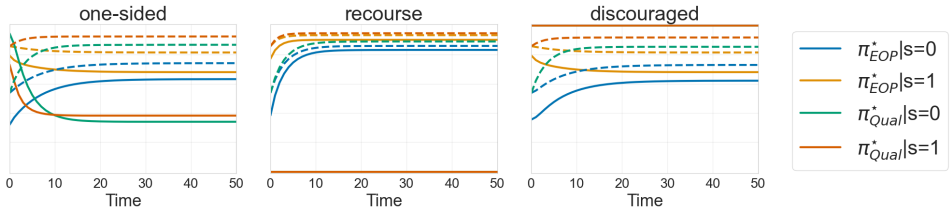


Figure 11: Loan probability $\mathbb{P}(D = 1 \mid S = s)$ (solid) and repayment probability $\mathbb{P}(Y = 1 \mid S = s)$ (dashed) for different type of dynamics (one-sided, recourse, discouraged) and policies $\pi^\star_{\text{EOP}}, \pi^\star_{\text{QUAL}}$ per sensitive attribute $s \in \{0, 1\}$. Initial distribution $\mu_0 = \text{FICO}$, $c = 0.8$, $\epsilon = 0.01$.

the `one-sided` dynamics, for $\pi^\star_{\text{QUAL}}$ the repayment rate is higher compared to the loan granting rate, indicating that a significant number of individuals who would repay their loan are not being granted one. In the case of `one-sided` dynamics, similar to the `discouraged` dynamics, we observe different short-term and long-term effects. Specifically, for $\pi^\star_{\text{EOP}}$, the probability of receiving a loan initially differs between the sensitive groups within the first 20 time steps. However, as time progresses, these probabilities tend to become closer to each other. This suggests a potential reduction in the disparity of loan access between the sensitive groups over time under the influence of the $\pi^\star_{\text{EOP}}$ policy. In the case of `recourse` dynamics, we observe that the loan granting and repayment probabilities tend to stabilize closely together in the long term across sensitive groups and under both policies—except for $\pi^\star_{\text{QUAL}}$ when $S = 1$. In this particular case, the $\pi^\star_{\text{QUAL}}$ policy sets $\pi(D = 1 \mid X = x, S = 1) = 0$ for all values of x. This scenario serves as an example where optimizing for long-term distributional goals without enforcing predictive fairness constraints can lead to individuals with a high probability of repayment being consistently denied loans.

E.5 Different Dynamic Speeds

We begin by assuming one-sided dynamics and then introduce variation in the speed of transitioning between different credit classes. This variation encompasses three levels: `slow`, `medium`, and `fast`, each representing the rate at which borrowers' credit scores evolve in response to decisions. Additional information about these specific dynamics can be found in Section D.2. For each of these three dynamics, we address both optimization problems. Solving (5) yields $\pi^\star_{\text{EOP}}$, while solving (6) provides $\pi^\star_{\text{QUAL}}$."

Feature and Outcome Trajectories. Figure 12 depicts the trajectories over 200 time steps for $\pi^\star_{\text{EOP}}$ and $\pi^\star_{\text{QUAL}}$ under different speeds of one-sided dynamics. While the initial distribution remains the same for all runs, the convergence process, time, and final stationary distribution ($\star$) vary depending on the dynamics speed. Regarding the group-dependent distribution of Y, we observe that $\pi^\star_{\text{QUAL}}$ achieves a higher distribution (which in addition is closer to the equal outcome distribution) compared to $\pi^\star_{\text{EOP}}$. This can be attributed to the fact that $\pi^\star_{\text{QUAL}}$ explicitly optimizes for maximizing the total distribution of Y. Additionally, we notice that for both policies slower dynamics result in lower stationary distributions of Y compared to faster dynamics.

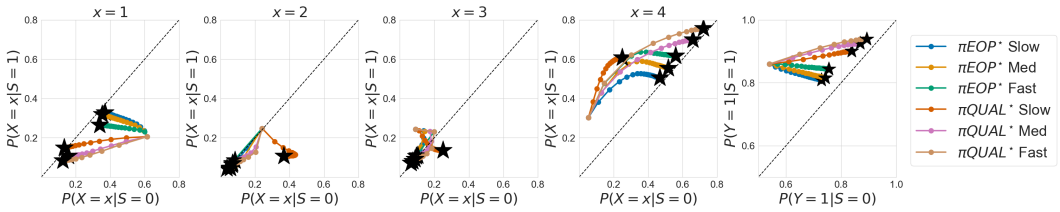


Figure 12: Convergence of $\pi^\star_{\texttt{EOP}}$ and $\pi^\star_{\texttt{QUAL}}$ for different speeds of dynamics towards different unique stationary distributions $\mu = \star$. Trajectories over 200 time steps. Left four plots: feature distribution μ_t. Right: distribution of the outcome of interest. Equal feature/outcome distribution dashed. Initial distribution μ_0 =FICO, $c = 0.8$, $\epsilon = 0.01$.

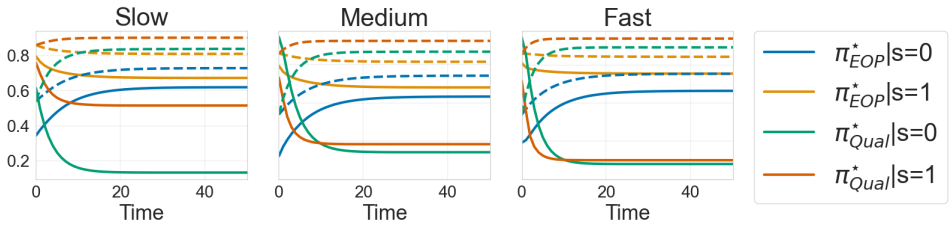


Figure 13: Loan probability $\mathbb{P}(D = 1 \mid S = s)$ (solid) and repayment probability $\mathbb{P}(Y = 1 \mid S = s)$ (dashed) for different speed of one-sided dynamics (`slow`, `medium`, `fast`) and policies $\pi^\star_{\texttt{EOP}}, \pi^\star_{\texttt{QUAL}}$ per sensitive attribute $s \in \{0, 1\}$. Initial distribution $\mu_0 =$ FICO, $c = 0.8$, $\epsilon = 0.01$.

Utility, Fairness and Loan and Repayment Probabilities. Figure 13 depicts the group-dependent probabilities of receiving a loan, $\mathbb{P}(D = 1 \mid S = s)$, and repayment, $\mathbb{P}(Y = 1 \mid S = s)$, for both non-privileged ($S = 0$) and privileged ($S = 1$) groups. The probabilities are shown for the convergence phase (initial 50 time steps) of policies $\pi^\star_{\texttt{EOP}}$ and $\pi^\star_{\texttt{QUAL}}$ across different speeds of one-sided dynamics. Higher payback probabilities compared to loan probabilities can indicate an underserved community where fewer credits are granted than would be repaid. Across all dynamics, we observe small differences in the repayment distributions for each policy. The repayment probabilities are consistently higher for the non-protected group compared to the protected group. Moreover, in general, $\pi^\star_{\texttt{QUAL}}$ yields higher repayment rates than $\pi^\star_{\texttt{EOP}}$. However, the loan probabilities—which indicate a group's access to credit—exhibit differences across dynamics and policies. As expected, the utility-maximizing $\pi^\star_{\texttt{EOP}}$ generally provides higher loan rates compared to $\pi^\star_{\texttt{QUAL}}$. While the loan rates remain similar across dynamics for $\pi^\star_{\texttt{EOP}}$, they vary for $\pi^\star_{\texttt{QUAL}}$. Under `slow` dynamics, $\pi^\star_{\texttt{QUAL}}$ yields low loan probabilities for the protected group, which then increases for `medium` and `fast` dynamics. Furthermore for $\pi^\star_{\texttt{QUAL}}$, the discrepancy between acceptance rates for sensitive groups is greatest at `slow` dynamics, and decreases significantly at `medium` dynamics - at the expense of the non-protected group. Finally, for `fast` dynamics, the acceptance rates for sensitive groups are approximately equal.

These observations emphasize the importance of conducting further investigations into the formulation of long-term goals, taking into account their dependence on dynamics and the short-term consequences. This includes not only considering the type of dynamics (one-sided or two-sided), but also the speed at which individuals' feature changes in response to a decision.

Effective Utility, Inequity and Unfairness. Figure 14 illustrates effective (accumulated) measures of utility, inequity, and (EOP) unfairness over time. For all dynamics, the policies align with their respective targets. $\pi^\star_{\texttt{EOP}}$ accumulates the highest utility across all dynamics while maintaining a low effective unfairness after an initial convergence period. On the other hand, $\pi^\star_{\texttt{QUAL}}$ exhibits a small negative effective utility due to the imposed zero-utility constraint, but achieves lower effective inequity by maximizing the total distribution of the outcome of interest. We observe that the speed of dynamics does not significantly affect effective utility for both policies and effective unfairness for the $\pi^\star_{\texttt{EOP}}$ policy. However the speed of dynamics does have an impact effective inequity, although its effect varies for each policy. Among the $\pi^\star_{\texttt{EOP}}$ policies, we find that the `medium` dynamics result

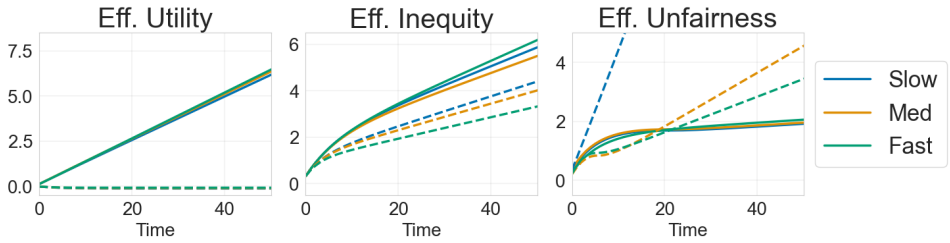
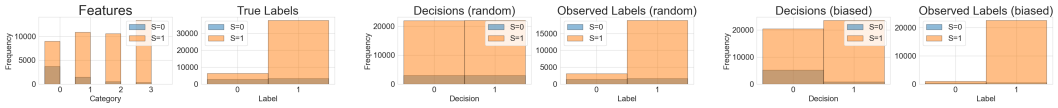


Figure 14: Effective (cumulative) utility $\mathcal{U}$, inequity $\mathcal{I}$, and (EOP) unfairness `EOPUnf` for different policies ($\pi^{\star}_{\texttt{EOP}}$ solid, $\pi^{\star}_{\texttt{QUAL}}$ dashed).



(a) True distributions of features and labels.

(b) Distribution of decisions and observed labels for `random`.

(c) Distribution of decisions and observed labels for `bias`.

Figure 15: Data distributions for different temporal datasets based on FICO used to estimate label distributions and dynamics.

in the lowest effective inequity, whereas among the $\pi^{\star}_{\texttt{QUAL}}$ policies, the `fast` dynamics exhibit the lowest effective inequity. While the effective utility is minimally affected by the speed of dynamics in the case of $\pi^{\star}_{\texttt{EOP}}$, we observe different results for effective inequity. Among the $\pi^{\star}_{\texttt{EOP}}$ policies, the `medium` dynamics result in the lowest effective inequity. Conversely, among the $\pi^{\star}_{\texttt{QUAL}}$ policies, the `fast` dynamics exhibit the lowest effective inequity. These observations highlight that the final outcomes of decision policies are not only influenced by the type of dynamics (one-sided and two-sided), but also by the speed of dynamics. It is thus crucial to also consider the rate at which individuals are able to change features within one time step. This consideration can for example be important in the context of recourse, where not all individuals may have the ability to implement the minimum recommended actions, potentially due to individual limitations. Consequently, only a fraction of individuals would be able to move up in their credit class in response to a negative decision.

E.6 Dynamics Estimation under Partially Observed Labels

We conduct additional experiments to investigate the impact of estimation errors in the underlying distributions on the quality of results. In a more realistic loan example, label Y might be partially observed (i.e., observed only for individuals who received a positive loan decision). In this case, the estimate of Y may no longer be as accurate for one sensitive group as for another. We investigate the sensitivity of our derived policy to the estimation of Y for different decision policies (which reveal different amounts of labels for different subgroups) compared to access to the true distribution of Y. We first generate a temporal dataset comprising two time steps. These samples were drawn from the FICO base distribution, and we assumed the dynamics of One-sided General (as described in § D.2). The dataset is comprised of 50,000 samples aligning with the dataset scales employed in the fairness literature, such as the Adult dataset Kohavi & Becker (2013). We deploy three different policies that influence the data observed at $t = 1$, `random`, `threshold`, `biased`, with the following formulations:

- `random` is defined by $\mathbb{P}(D = 1 \mid X, S) = 0.5$ for all X, S;
- `bias` is defined for all S by $\mathbb{P}(D = 1 \mid X, S) = 0.1$ if $X <= 2$ and for $S = 0$ as $\mathbb{P}(D = 1 \mid X, S) = 0.3$ if $X > 2$ and for $S = 1$ as $\mathbb{P}(D = 1 \mid X, S) = 0.9$.

The true distribution of features and label at $t = 0$ are shown in Figure 15a. The distributions of decisions and observed labels under the different policies are shown in Figures 15b - 15c.

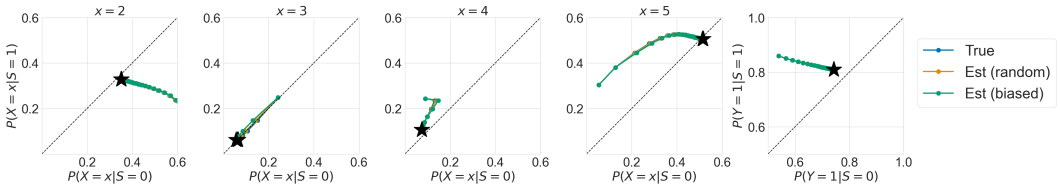


Figure 16: Convergence of $\pi^{\star}_{\texttt{EOP}}$ under true and estimations of $\ell(y \mid x, s)$ and $g(k \mid x, d, y, s)$ and under different type of initial policies (`random`, `threshold`, `bias`). 200 time steps, last time step marked $\star$. Top four plots: feature distribution μ_t. Bottom left: distribution of the outcome of interest. Equal feature/outcome distribution dashed.

We then estimate both $\ell(y \mid x, s)$ and $g(k \mid x, d, y, s)$ from the observed samples, with the latter being dependent on the former. Subsequently, we solve the optimization problem ($c = 0.9$, $\epsilon = 0.00005$) using these estimated distributions yielding three different policies (one per estimation). Consequently, we simulate the performance of the discovered policies under the true distributions and μ_0 =FICO. In the evaluation, we compare the results to the policy obtained under the true probability estimate $\ell(y \mid x, s)$ as supplied by FICO (`true`).

Feature and Outcome Trajectories. Figure 16 displays the trajectories of $\pi^{\star}_{\texttt{EOP}}$ for 200 time steps for the optimal policies obtained under both the true and estimated distributions and dynamics. Notably, the initial distribution remains the same, and the policies slightly vary in their convergence process to the stationary distribution ($\star$), while staying close to each other. It is important to emphasize that all policies successfully achieve a stationary distribution. This is due to the fact that even though we employ estimated distributions as inputs for the optimization problem, we are still solving the optimization problem for a policy that induces a stationary distribution that satisfies the fairness criteria. We showcase this in the next results.

Utility, Fairness and Loan and Repayment Probabilities. Figure 17 (left) displays $\mathcal{U}$ and `EOPUnf` over the first 50 time steps (until convergence). We observe that the policies exhibit a different level of unfairness, while still achieving low unfairness. The policy derived from the true probabilities and dynamics achieves lowest unfairness, the policy derived from probabilities and dynamics collected under a random policy has slightly higher unfairness, and the policy derived from probabilities and dynamics collected under a biased policy has the highest unfairness. In terms of utility, where we aim for maximization without imposing a strict constraint, we observe that all policies exhibit a similar utility level. Figure 9 (middle, right) displays the loan probability $\mathbb{P}(D = 1 \mid S = s)$ and payback probability $\mathbb{P}(Y = 1 \mid S = s)$ for non-privileged ($S = 0$) and privileged ($S = 1$) groups. While there is no difference in loan and payback probabilities for the privileged group ($S = 1$) between the policies, we observe a small difference for the unprivileged group ($S = 0$). The policy derived from true probabilities and dynamics provides fewer loans to the unprivileged group compared to the policy derived from probabilities and dynamics collected under the random policy. Interestingly, the policy derived from probabilities and dynamics collected under a biased policy grants the most loans to the unprivileged group. Note, that our unfairness metric in the left plot is equal opportunity Hardt et al. (2016b), not demographic parity Dwork et al. (2012). Consequently, this observation may be explained by the policy obtained from biased estimation providing loans to a higher number of individuals from the unprivileged group who may not be able to repay them. Thus, while we do achieve a stationary distribution using estimated probabilities, it is important to note that convergence to the intended fair state is not guaranteed when estimation errors are present. However, if the estimations closely approximate the true distribution, the resulting stationary distribution achieves similar utility and fairness properties as the stationary distribution that would have been achieved had the policy found under the true probabilities.

F Example Scenarios

F.1 Assumptions of the Guiding Example

In this section, we discuss the assumptions taken in the data generative model introduced in § 2.

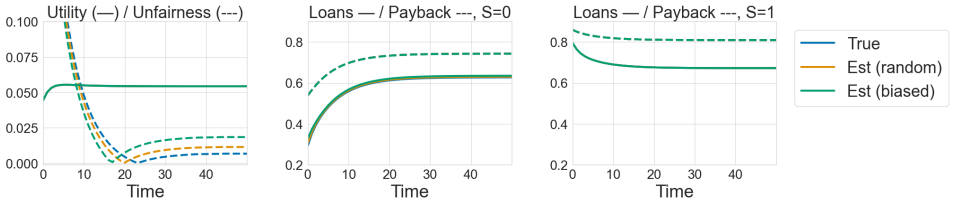


Figure 17: Results for our $\pi^{\star}_{\texttt{EOP}}$ under true and estimations of $\ell(y \mid x, s)$ under different type of initial policies (`random`, `threshold`, `bias`). Top Left: Utility (solid, ↑) and EOP-Unfairness (dashed, ↓) over first 50 time steps. Remaining: Loan (solid) and payback probability (dashed) per policy and sensitive S.

Assumptions F.1. *S is a root node and X_t, Y_t and D_t (potentially) depend on S.*

It is commonly assumed in the causality and fairness literature that sensitive features are root nodes in the graphical representation of the data generative model (Kusner et al., 2017a; Chiappa, 2019; Kilbertus et al., 2020a), although there is some debate on this topic (Mhasawade & Chunara, 2021; Hu & Kohler-Hausmann, 2020). The assumption that the sensitive attribute S influences X_t is based on the observation that in practical scenarios, nearly every (human) characteristic is causally influenced by the sensitive attribute (Kusner et al., 2017a; Chiappa, 2019). In some cases, it is also assumed that S influences Y_t (Chiappa, 2019), while in other cases, this assumption is not made (Liu et al., 2018). The extent to which the decision D_t is directly influenced by the sensitive attribute S depends on the decision policy being employed. Policies that strive for (statistical) fairness often require explicit consideration of the protected attribute in their decision-making process (Hardt et al., 2016b; Dwork et al., 2012; Corbett-Davies et al., 2017).

Assumptions F.2. *The outcome of interest Y_t depends on features X_t.*

The assumption that changes in X_t lead to changes in Y_t is prevalent in scenarios involving lending (Liu et al., 2018; Creager et al., 2020; D'Amour et al., 2020; Hu & Zhang, 2022). This assumption is also implicit in problems where individuals seek recourse, e.g., via minimal consequential recommendations (Karimi et al., 2021a) or social learning (Heidari et al., 2019).

Assumptions F.3. *Decision D_t depends on features X_t.*

In algorithmic decision-making, the primary objective of a policy is typically to predict the unobserved label or outcome of interest, denoted as Y, based on the observable features, denoted as X (Schölkopf et al., 2012). We make the assumption that an individual's observed features at a particular point in time are sufficient to make a decision and conditioned on these features, the decision is independent of past features, labels, and decisions. This assumption aligns with prior work in the field (Zhang et al., 2020; Creager et al., 2020; Karimi et al., 2021a).

Assumptions F.4. *An individual's sensitive attribute S is immutable over time.*

For simplicity, we assume that individuals do not change their sensitive attribute. This assumption aligns with previous works that consider a closed population (Liu et al., 2018; Creager et al., 2020; D'Amour et al., 2020; von Kügelgen et al., 2022). A closed population refers to a group of individuals that remains constant throughout the study or analysis. It implies that there are no additions or removals from the population of interest. Other work considers that individuals join and leave the population over time, leading to a changing distribution of the sensitive attribute (Hashimoto et al., 2018). The assumption that individuals do not change their sensitive attribute is controversial because, on the one hand, social categories are often ontologically unstable (Barocas et al., 2019; Hu & Kohler-Hausmann, 2020), and as such their boundaries are not clearly defined and dynamic. On the other hand, it ignores that individuals may be assigned identities at birth which they have the agency to correct at a given time. For example, an individual assigned one religion at birth may have a different religion at a later stage in life.

Assumptions F.5. *An individual's next step's features X_{t+1} depend on its current step's feature X_t, decision D_t, outcome of interest Y_t, and sensitive S.*

This assumption, as discussed in previous literature, can be attributed to either bureaucratic policies (Liu et al., 2018) or changes in individual behavior, in response to recommendations (Karimi

et al., 2021b) or social learning (Heidari et al., 2019). In the lending context, it is commonly assumed that the higher the credit score the better. Then the assumption is: individuals approved for a loan ($D = 1$) experience a positive score change upon successful repayment ($Y = 1$) and a negative score change in case of default ($Y = 0$), while individuals rejected for a loan ($D = 0$) are assumed to have no score change (Liu et al., 2018; Creager et al., 2020; D'Amour et al., 2020). In scenarios where individuals who are not granted a loan ($D = 0$) seek recourse, it would be assumed that a negative decision leads to an increase in credit score, to elicit a positive decision change in subsequent time steps (Heidari et al., 2019; Karimi et al., 2021b).

For the transition probabilities to be time-homogeneous, we take the following assumptions:

Assumptions F.6. *Dynamics $g(k \mid x, d, y, s)$ remain fixed over time.*

This is a common assumption in the literature (Zhang et al., 2020; Creager et al., 2020; D'Amour et al., 2020; Hu & Zhang, 2022). Although real-world data often exhibits temporal changes, we make the simplifying assumption of static dynamics. We can treat the dynamics as constant for specific durations. This is reasonable in situations where changes are based on policies involving bureaucratic adjustments (Liu et al., 2018) or algorithmic recourse recommendations (Karimi et al., 2022), and where it is desirable for these policies to remain unchanged or not be retrained at every time step (Perdomo et al., 2020). In practical applications, MDPs with time-varying transition probabilities present challenges, and the literature addresses this through online learning algorithms (e.g., (Yu & Mannor, 2009; Li et al., 2019)).

Assumptions F.7. *Label distribution $\ell(y \mid x, s)$ remains fixed over time.*

This assumption is widely recognized in the literature (Heidari et al., 2019; Zhang et al., 2020; Creager et al., 2020; D'Amour et al., 2020; Karimi et al., 2021b; Hu & Zhang, 2022). However, in real-world scenarios, the relationship between input data X_t and the target output Y_t may change over time, resulting in changes in the conditional distribution $\ell(y \mid x, s)$. This phenomenon is commonly referred to as *concept drift* (Lu et al., 2018; Gama et al., 2014). In the lending scenario, concept drift may arise from changes in individuals' repayment behavior or alterations in the process of generating credit scores based on underlying features like income, assets, etc.

F.2 Additional Example: Qualifications over Time

In this section, we provide an additional example, which could also be covered by our framework. The example was provided by (Zhang et al., 2020) with their data generative model displayed in Figure 18. The primary distinction from the example presented in Section 2 lies in the assumption that $Y_t \to X_t$. (Zhang et al., 2020) employ their model to replicate lending and recidivism scenarios over time in their experiments, using FICO and COMPAS data, respectively. However, most prior work has modeled the (FICO) lending examples as $X_t \to Y_t$ (Liu et al., 2018; Creager et al., 2020; D'Amour et al., 2020). The same holds for recidivism (COMPAS) (Russell et al., 2017). We, therefore, frame the example as a repeated admission example where Y_t denotes a (presumably hidden) qualification state at time t, following (Rateike et al., 2022b; Kusner et al., 2017b).

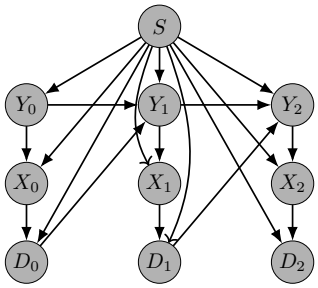


Figure 18: Data generative model. Time steps (subscript) $t = \{0, 1, 2\}$.

Data Generative Model. Let an individual with protected attribute S (e.g., gender) at time t be described by a qualification Y_t and a non-sensitive feature X_t (e.g., grade or recommendations levels). We assume the sensitive attribute to remain fixed over time, and drop the attributes time subscript. For simplicity, we assume binary sensitive attribute and qualification, i.e., $S, Y_t \in \{0, 1\}$ and a one-dimensional discrete non-sensitive feature $X_t \in \mathbb{Z}$. Let the population's sensitive attributes be distributed as $\gamma(s) := \mathbb{P}(S = s)$ and assume them to remain constant over time. We assume Y_t to depend on S, such that the group-conditional qualification distribution at time t is $\mu_t(y \mid s) := \mathbb{P}(Y_t = y \mid S = s)$. For example, different demographic groups may have different qualification distributions due to structural discrimination in society. We assume that the non-sensitive features X_t are influenced by the qualification Y_t and, possibly (e.g.,

due to structural discrimination), the sensitive attribute S. This leads to the feature distribution $f(x \mid y, s) := \mathbb{P}(X_t = x \mid Y_t = y, S = s)$, We assume that there exists a policy that takes at each time step t binary decisions D_t (e.g., whether to admit) based on X_t and (potentially) S and decides with probability $\pi(d \mid x, s) := \mathbb{P}(D_t = d \mid X_t = x, S = s)$.

Consider now dynamics in which the decision D_t made at one time step t, directly impacts an individual's qualifications at the next step, Y_{t+1}. Assume the transition from the current qualification state Y_t to the next state Y_{t+1} is determined by the current qualification state Y_t, decision D_t and (potentially) sensitive attribute S. For example, upon receiving a positive admission decision, an individual may be very motivated and increase their qualifications. However, due to structural discrimination, the extent of the qualification change may be influenced by the individual's sensitive attribute. We denote the probability of an individual with $S = s$ changing from qualification $Y_t = y$ to $Y_{t+1} = k$ in the next step in response to decision $D_t = d$ as dynamics $g(k \mid y, d, s) := \mathbb{P}(Y_{t+1} = k \mid Y_t = y, D_t = d, S = s)$. Crucially, the next step qualification state (conditioned on the sensitive attribute) depends only on the present state qualification and decision, and not on any past states.

Dynamical System. We can now describe the evolution of the group-conditional qualification distribution $\mu_t(y \mid s)$ over time t. The probability of a qualification change from y to k in the next step given s is obtained by marginalizing out decision D_t, resulting in

$$\mathbb{P}(Y_{t+1} = k \mid Y_t = y, S = s) = \sum_{xd} g(k \mid y, d, s)\pi(d \mid x, s)f(x \mid y, s). \tag{36}$$

These transition probabilities together with the initial distribution over states $\mu_0(y \mid s)$ define the behavior of the dynamical system. In our model, we assume that the dynamics $g(k \mid y, d, s)$ are time-independent, meaning that the qualification changes in response to the decision, the previous qualification and the sensitive attribute remain constant over time. We also assume that the distribution of the non-sensitive features conditioned on an individual's qualification and sensitive attribute $f(x \mid y, s)$ does not change over time (e.g., individuals need a certain qualification to generate certain non-sensitive features). Additionally, we assume that the policy $\pi(d \mid x, s)$ can be chosen by a policy maker and may depend on time. Under these assumptions, the probability of a feature change depends solely on the policy π and sensitive feature S.