Probabilistic Approaches for Fair Clustering

Abhisek Chakraborty
Department of Statistics, Texas A&M University

Joint work with Anirban Bhattacharya and Debdeep Pati

March 4, 2024
Stat Cafe

Modules

1. Fairness in Clustering
2. Hierarchical Fair-Dirichlet Process
3. Generalised Bayesian Clustering

Disparate Impact Doctrine and Clustering

- The rise of machine learning driven decision-making sparked a growing emphasis on algorithmic fairness.
- Chierichetti et al. (2017) introduced the notion of fairness in clustering .
- Disparate impact doctrine (Feldman et al., 2015) dictates that the decisions made should not be disproportionately different for individuals belonging to different labels of Protected attribute.
- **A Silly Example.** Height can be closely tied to a protected feature like gender. Decisions based on height may unfairly discriminate.

Balance

- Observe data $\{(\boldsymbol{x}_i, a_i)\}_{i=1}^N$, with $\boldsymbol{x}_i$ denoting d-variate features, and $a_i \in \{1, \ldots, r\}$ denoting the label of the protected attribute.
- **Goal.** Assign the data points into clusters $\boldsymbol{C} = (C_1, \ldots, C_K)$, respecting balance.

Definition (Balance, (Chierichetti et al., 2017))

The balance in C_k is defined as

$$\text{Balance}(C_k) = \min_{1 \le j_1 < j_2 \le r} \min \left\{ \frac{|C_{kj_1}|}{|C_{kj_2}|}, \frac{|C_{kj_2}|}{|C_{kj_1}|} \right\}$$

where $|C_{kj}|$ denote the number of observations in C_k with $a = j$. The overall balance of the clustering is

$$\text{Balance}(\boldsymbol{C}) = \min_{k=1,\ldots,K} \text{Balance}(C_k)$$

The higher this measure is, the fairer is the clustering.

Fair Clustering via Fairlets

- Given the notion of balance, Chierichetti et al. (2017) introduced the concept of fairlets, i.e minimal fair sets that approximately maintain the selected clustering objective.
- **Fairlet Decomposition.** Any fair clustering problem involves initially obtaining a fairlet decomposition of the data through the solution of a minimum cost flow (NP-Hard) problem.
- **Clustering Fairlets.** Classical clustering algorithms, such as k-means or k-center, can be employed for further processing.

Minimum Cost Flow via Bipartite Graph Matching

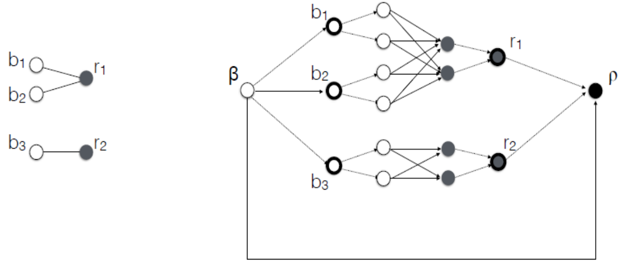


Figure 2: The construction of the MCF instance for the bipartite graph for $t' = 2$. Note that the only nodes with positive demands or supplies are $\beta, \rho, b_1, b_2, b_3, r_1$, and r_2 and all the dotted edges have cost 0.

Figure: Adopted from Chierichetti et al. (2017).

Related Works

- Multi-color case (BÃ¶hm et al., 2020); Imperfect knowledge of group membership (Esmaeili et al., 2020), etc.
- Fairness in other avatars of clustering, e.g spectral clustering (Kleindessner et al., 2019), correlation clustering (Ahmadian et al., 2020b), hierarchical clustering (Ahmadian et al., 2020a).
- Alternative notions of fairness in clustering, e.g individual fairness (Kleindessner et al., 2020; Mahabadi and Vakilian, 2020; Chakrabarty and Negahbani, 2021), proportional fairness (Chen et al., 2019).
- Fairness in clustering + Other pressing aspects of modern machine learning, e.g Privacy (Rösner and Schmidt, 2018), Robustness (Bandyapadhyay et al., 2019), etc.

Motivation for a probabilistic approach

- We take a novel generative model-based approach to tackle the problem of clustering under balance constraints.
- Measure of uncertainty? Probabilities of cluster assignments.
- Non-Gaussian data accommodated easily.
- Notion of the true fair clustering configuration at the population level?

Broader connections

- Interplay between probability modeling perspective and optimization perspective have a rich history in statistics.
- Quantile regression/asymmetric Laplace errors.
- PCA/probabilistic PCA
- k-means clustering/Gaussian mixture model.
- Current contribution aims to add a similar perspective in the context of fair clustering.

Notion of Balance at Population Level

- We assume that $\{(x_i, a_i)\}_{i=1}^N$ are independent copies of a random vector $(X, A) \in \mathcal{X} \times A \subset \mathbb{R}^d \times [r]$.
- A weight vector $\boldsymbol{\xi} \in \Delta^{r-1}$ records the population proportions of the different labels of the protected attribute.
- **Generative Model.**

  $$A \sim \mathcal{P}_A^\star \equiv \text{Multinomial}(1, \boldsymbol{\xi}), \ (X, Z) \mid A \sim \mathcal{P}_{X,Z|A}^\star \equiv \mathcal{P}_{X|Z,A}^\star \times \mathcal{P}_{Z|A}^\star,$$

  where $\mathcal{P}_{X,Z,A}^\star$ is unknown and Z is the latent/unobserved clustering index.
- We only observe independent copies of (X, A); learn the marginal generative mechanism $\mathcal{P}_Z^\star$ of the clustering index Z modulo fairness.

Notion of Balance at Population Level

- To define Balance at population level, consider

$$\mathbb{P}^{(R)} = \{\mathcal{P}_{X,Z,A} \ : \ \mathcal{P}_{A|Z} = \mathcal{P}_A\}$$

  to be all joint distributions such that every label of the protected attribute are equally likely to appear within each cluster.

- To allow some small departure, define

$$\mathbb{P}_\varepsilon^{(R)} = \{\mathcal{P}_{X,Z,A} \ : \ \text{KL}(\mathcal{P}_A \times \mathcal{P}_Z \,||\, \mathcal{P}_{A,Z}) \leq \varepsilon\},$$

  where $\varepsilon \geq 0$ controls the extent of departure from balance.

- $\text{KL}(\mathcal{P}_A \times \mathcal{P}_Z \,||\, \mathcal{P}_{A,Z})$ is the *mutual information* between (A, Z).

- Say that (A, Z) satisfies ε-*balance* under $\mathcal{P}_{X,Z,A}$ if $\text{KL}(\mathcal{P}_A \times \mathcal{P}_Z \,||\, \mathcal{P}_{A,Z}) \leq \varepsilon$.

Notion of Balance at Population Level

- The true generative model $\mathcal{P}^\star$ may not belong to $\mathbb{P}_\varepsilon^{(R)}$.
- Since we wish to ensure that our clustering procedure is ε-*balanced*, the inferential goal constitutes finding the “best" approximation of the true generative model $\mathcal{P}^\star$ within the restricted class $\mathbb{P}_\varepsilon^{(R)}$.
- Reminiscent of maximum likelihood estimation under model misspecification (White, 1982). Define as KL projection.
- Methodologically, enforce the constraint in a soft manner through the prior.

Fair Clustering via Hierarchical Fair Dirichlet Process

- Hierarchical Bayesian model to carry out clustering with fairness constraints (`arxiv.org/pdf/2305.17557`).
- Lowest level hyperparameters of our model are (K, g, b), where K is an upper bound on the number of clusters, and g, b are positive parameters.
- Let Δ^{K-1} be the $(K-1)$-dimensional probability simplex; i.e. all $(p_1, \ldots, p_K)$ with $p_i \geq 0$ for all i and $\sum_{i=1}^{K} p_i = 1$.

Some more notation

- Let $\mathcal{Z}_{N,K} = \{\boldsymbol{z} = (z_1, \ldots, z_N) : z_i \in [K] \text{ for all } i \in [N]\}$ denote the space of all clustering configurations of N observations into K clusters.

- Any $\boldsymbol{m} = (m_1, \ldots, m_K) \in \mathbb{Z}_{\geq 0}^K$ such that $\sum_{k=1}^K m_k = N$ called a *cluster occupancy vector*.

- Given such $\boldsymbol{m}$, let $\mathcal{Z}_{N,K,\boldsymbol{m}} = \{z \in \mathcal{Z}_{N,K} \ : \sum_{i=1}^N \mathbf{1}(z_i = k) = m_k, \ i \in [N]\}$ denote all clustering configurations with cluster occupancy vector $\boldsymbol{m}$.

- $\mathcal{Z}_{N,K,\boldsymbol{m}}$ can be uniquely characterized by the space of $N \times K$ binary cluster membership matrices with fixed column-sum $\boldsymbol{m}$ and row-sum $\mathbf{1}_N$.

HFDP

- Given (K, g, b), sample a global weight vector $\boldsymbol{\beta} \in \Delta^{K-1}$ and a concentration parameter $\alpha_0 \in (0, \infty)$

$$\boldsymbol{\beta} \mid g \sim \text{Dir}(g/K, \ldots, g/K); \quad \alpha_0 \mid g, b \;\sim\; \text{Gamma}(g,\; b).$$

- Next, given α_0 and $\boldsymbol{\beta}$, independently sample a weight vector $\boldsymbol{w}^{(a)}$ corresponding to each level of the attribute $a \in [r]$

$$\boldsymbol{w}^{(a)} \mid \alpha_0, \boldsymbol{\beta} \overset{\textit{ind.}}{\sim} \text{Dir}(\alpha_0\, \boldsymbol{\beta}),\; a \in [r].$$

- The concentration parameter α_0 dictates how tightly the $\{\boldsymbol{w}^{(a)}\}_{a=1}^{r}$ concentrate around $\boldsymbol{\beta}$.

- Critical in enabling a notion of balance in our model-prior specification.

HFDP

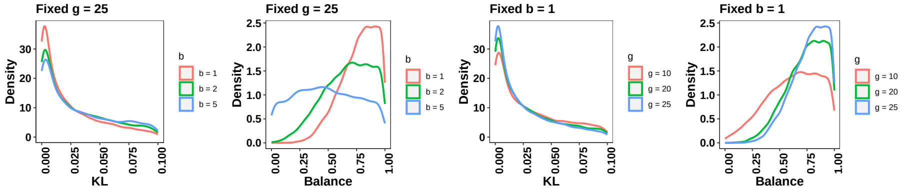


Figure: Prior calibration: Given (g, b) and $K = 2$, we obtain prior draws of $(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)})$ as above, and present the induced prior distribution of KL$(w^{(1)} \mid\mid w^{(2)})$ and balance between $(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)})$. The first two plots present the distribution of KL$(w^{(1)} \mid\mid w^{(2)})$, and balance between $(\boldsymbol{w}^{(1)}, \boldsymbol{w}^{(2)})$ respectively, for varying b with fixed g. The final two plots present the same quantities, now with varying g with fixed b.

HFDP

- Define a function rd : $\mathbb{N} \times \Delta^{t-1} \to \mathbb{Z}^t_{\geq 0}$, so that for a positive integer $n \in \mathbb{N}$ and a probability vector $\boldsymbol{u} \in \Delta^{t-1}$, $\boldsymbol{v} = \mathrm{rd}(n, \boldsymbol{u})$ is given by $v_i = \mathrm{round}(nu_i)$ for $i \in [t-1]$, where “round" denotes the rounding function to the nearest integer, and $v_t = n - \sum_{i=1}^{t-1} v_i \geq 0$.
- Clearly, $\langle 1_t, \mathrm{rd}(n, \boldsymbol{u}) \rangle = n$ for any $\boldsymbol{u}$. We use this rd$(\cdot)$ function to create a novel prior on cluster occupancy vectors.
- Having drawn $\boldsymbol{w}^{(a)}$ for each label of the attribute, set

$$\boldsymbol{m}^{(a)} = \mathrm{rd}\big(N^{(a)}, \boldsymbol{w}^{(a)}\big), \quad a \in [r].$$

- Call the induced prior on $\boldsymbol{m}^{(a)}$ a *lifted Dirichlet* prior with parameters $N^{(a)}, \alpha_0, \beta$.

HFDP

- Dirichlet-Multinomial prior more commonly used in the literature, where $\boldsymbol{m}^{(a)}$ is additionally *sampled* from a Multinomial distribution with total count $N^{(a)}$ and probability vector $\boldsymbol{w}^{(a)}$.
- Instead, the randomness in a lifted Dirichlet prior is entirely controlled by $\boldsymbol{w}^{(a)}$, enabling tighter control on the cluster sizes across a.
- This is crucial towards enforcing balance *a priori* in our framework.

HFDP (contd.)

- Recap: hierarchical specification so far

$$\boldsymbol{m}^{(a)} = \mathsf{rd}(N^{(a)}, \boldsymbol{w}^{(a)}),\ \boldsymbol{w}^{(a)} \,|\, \alpha_0, \boldsymbol{\beta} \overset{\text{ind.}}{\sim} \mathsf{Dir}(\alpha_0\, \boldsymbol{\beta}), \quad a \in [r],$$
$$\boldsymbol{\beta} \,|\, g \sim \mathsf{Dir}(g/K, \ldots, g/K); \quad \alpha_0 \,|\, g, b \sim \mathsf{Gamma}(g,\ b).$$

- Having drawn $\boldsymbol{m}^{(a)}$, draw the cluster configuration $\boldsymbol{z}^{(a)}$

$$\boldsymbol{z}^{(a)} \mid \boldsymbol{m}^{(a)} \overset{ind.}{\sim} \mathsf{Unif}(\mathcal{Z}_{N_a, K, \boldsymbol{m}^{(a)}}), \quad a \in [r].$$

- A probabilistic clustering mechanism subject to fairness constraints.
- Can be embedded into any probability model for data within clusters.

HFDP (contd.)

- Specifying component-wise distributions:

$$\boldsymbol{x}_i^{(a)} \mid z_i^{(a)} = k, \boldsymbol{\phi}_k^{(a)} \overset{\textit{ind.}}{\sim} f_{obs}(\cdot \mid \boldsymbol{\phi}_k^{(a)}), \quad i \in [N_a],\ a \in [r]$$
$$\boldsymbol{\phi}_k^{(a)} \mid \boldsymbol{\phi}^{(a)} \overset{\textit{ind.}}{\sim} f_{pop}(\cdot \mid \boldsymbol{\phi}^{(a)}), \quad k \in [K],\ a \in [r],$$
$$\boldsymbol{\phi}^{(a)} \overset{\textit{i.i.d}}{\sim} f_{atom}, \quad a \in [r],$$

- For example, for a Gaussian model for continuous data, $f_{obs}(\cdot \mid \boldsymbol{\phi}_k^{(a)}) = \mathsf{N}_d(\boldsymbol{\mu}_k^{(a)},\ \Sigma_k^{(a)})$ where $\phi_k^{(a)} = (\boldsymbol{\mu}_k^{(a)},\ \Sigma_k^{(a)})$.

- Non-Gaussian models easily incorporated.

Posterior Computation

- **Sampling from** $[\beta \mid \cdot]$ **and** $[\alpha_0 \mid \cdot]$
  The prior $\alpha_0 \,|\, g, b \;\sim\;$ Gamma$(g,\ b)$ reduces the problem to sampling from log-concave densities. A simple rejection sampler with a well-designed covering density works.

- **Sampling from** $[\boldsymbol{w}^{(a)} \mid \cdot],\ a \in [r]$
  Admits closed from updates.

- **Sampling from** $[\boldsymbol{z}^{(a)} \mid \cdot],\ a \in [r]$
  Major computational bottleneck! Crucial utilization of Optimal Transport and a novel weighted sampling scheme in the space of binary matrices with fixed margins.

- To summarize the MCMC output and obtain a point estimate for the fair clustering configuration, we adopt the least-squares model-based clustering method of Dahl (2006).

Posterior Computation

- The marginal conditional of clustering indices $[\boldsymbol{z}^{(a)} \mid -],\ a \in [r]$, integrating out population parameters, is

$$[\boldsymbol{z}^{(a)} \mid -] \ \propto \ \prod_{k=1}^{K} \frac{\Gamma_d(\nu_k^{(a)}/2)\ (\lambda_0^{(a)})^{d/2}\ |\Lambda_0^{(a)}|^{\nu_0^{(a)}/2}}{\Gamma_d(\nu_0^{(a)}/2)\ (\lambda_k^{(a)})^{d/2}\ |\Lambda_k^{(a)}|^{\nu_k^{(a)}/2}}, \quad \boldsymbol{z}^{(a)} \in \mathcal{Z}_{N_a, K, \boldsymbol{m}^{(a)}}$$

- Difficult combinatorial problem and presents the most substantial computational challenge in our algorithm.

- Recast the problem as a non-uniform sampling task from the space of binary matrices with fixed margins (Miller and Harrison, 2013; Wang, 2020).

- Specifically, adapt rectangular loop algorithm of Wang (2020) to the weighted setting.

Experiment (Well-specified)

- Generate data with two attributes and two clusters.
- First, 20 individuals with $a = 1$ are generated from $N_2(\mu_{11}, S)$ and 30 individuals with $a = 2$ are generated from $N_2(\mu_{21}, S)$.
- Next, 30 individuals with $a = 1$ are generated from $N_2(\mu_{12}, S)$ and 20 individuals with $a = 2$ are generated from $N_2(\mu_{22}, S)$.
- Here, $\mu_{11} = (4, 4)',\ \mu_{21} = (2, 2)',\ \mu_{12} = (10, 10)',\ \mu_{22} = (8, 8)'$, and $S = 3 \times [\rho 11^{\mathrm{T}} + (1 - \rho) I_2]$ with $\rho = 0.3$.
- This generating mechanism ensures that individuals with $a = 1$ and $a = 2$ are equally represented in the observed sample of size 100. The goal is to obtain 2 balanced clusters.

Experiments (Mis-specified, Multi-color)

Mis-specified case. We follow the same scheme as before, except for simulating from multivariate t-distributions with centers $(\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22})$, scale S, and degrees of freedom 4. For multivariate skew normal distributions with centers $(\mu_{11}, \mu_{21}, \mu_{12}, \mu_{22})$, scale S, and the skewness parameter $\alpha = (1, 1)^{\mathrm{T}}$.

Multi-color case. Number of colors $r = 4$, sample sizes $N^{(1)} = N^{(2)} = N^{(3)} = N^{(4)} = 200$, true number of clusters $K_{true} = 2$. Follow a scheme similar to the Well specified case earlier.

Experiment

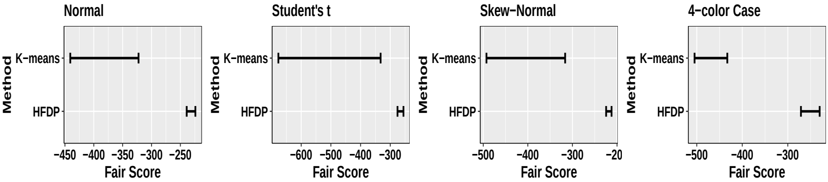


Figure: *Two/Multi-color case. The MAP of the HFDP fares better compared to fair-clustering with fairlets (termed as K-Means) in two color case, and the method in BÃ¶hm et al. (2020) (termed as K-Means) in multi-color case.*

.

Benchmark Datasets

- Compare methods on popular bench mark data sets from the UCI repository.
- **Datasets.** (1) Diabetes data (Variables: age, time in hospital; Protected attribute: gender). (2) Portuguese Banking data (Variables: age, balance, and duration; Protected attribute: marital status). (3) Credit card data (Variables: age, credit limit; Protected attribute: marital status).

Dataset	Attributes	$\hat{K}$	HFDP	Fairlet
Diabetes	2	5	-114.74	-651.84
Portuguese Banking	2	5	-176.85	-639.12
Credit Card	3	3	-86.19	-271.09

What did we discuss so far?

- **Existing Literature.** Uncertainty quantification was largely illusive .
- **HFDP.** Proposed a model-based approach. Developed a concrete notion of optimal recovery and principled performance evaluation.
- **Limitations of HFDP.** Model-based fair clustering frameworks show brittleness under model-misspecification and can be computationally prohibitive.
- **Remedy.** To circumnavigate such issues, we next propose a generalised Bayesian fair clustering framework that inherently enjoys decision theoretic interpretation, and support efficient computation.

Generalised Bayesian Posterior

- Let $\mathbf{u} = (u_1, \ldots, u_N)^{\mathrm{T}}$ be the observed data, $\theta \in \Theta$ is the parameter. Loss function minimization:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta \mid \mathbf{u}).$$

- Assume prior $\pi(\theta)$ on θ. Gibbs posterior:

$$\pi(\theta \mid \lambda, \mathbf{u}) \propto \pi(\theta) \exp\{-\lambda \mathcal{L}(\theta \mid \mathbf{u})\},$$

  where $\lambda > 0$ is a temperature parameter. Standard Bayesian inference recovered with $\lambda \mathcal{L}(\theta \mid \mathbf{u})$ as negative log-likelihood.

- Gibbs posterior offers a rational update of beliefs (Bissiri et al., 2016). Holmes and Walker (2017) proposed scheme for tuning λ.

Loss Functions

Fairlet Decomposition

- Given data $\{(\boldsymbol{x}_i, a_i) \in \mathcal{X} \times [2],\ i \in [N]\}$, Chierichetti et al. (2017) involves first decomposing data into a set of m fairlets, and calculate the m fairlet centers.
- Let $\mathcal{L}_1 : \mathcal{U} \to \mathbf{R}^+$ denote loss function for the fairlet decomposition.

Clustering the Fairlet Centers

- Factorized loss:

$$\mathcal{L}_2(\mathbf{C} \mid \mathbf{u}^\star) = \sum_{k=1}^{K} \sum_{i \in C_k} \mathcal{D}(u_i^\star, \mathbf{u}_k^\star), \qquad \mathbf{C} : |\mathbf{C}| = K,$$

  where $\mathcal{D}(u_i^\star, \mathbf{u}_k^\star) \geq 0$ is the discrepancy between $u_i^\star$ and $\mathbf{u}_k^\star$.
- K-means loss: $\mathcal{L}_2(\mathbf{C} \mid \mathbf{u}^\star) = \sum_{k=1}^{K} \sum_{i \in C_k} ||u_i^\star - \mathbf{u}_k^\star||_2^2$.

Gibbs Posterior for Fair Clustering

Priors

- $\mathcal{U}(\subset \mathcal{X}^m)$ denote the class of all "m fairlet centers". Uniform prior on $\mathcal{U}$.
- Uniform clustering priors.

Posterior

- $\pi(\mathbf{C}, \mathbf{u} \mid (\lambda_1, \lambda_2), \{(\boldsymbol{x}_i, a_i)\}_{i=1}^N) \propto$

$$\frac{\exp\{-\lambda_1 \mathcal{L}_1(\mathbf{u})\}}{\sum_{\mathbf{u}\in\mathcal{U}} \exp\{-\lambda_1 \mathcal{L}_1(\mathbf{u})\}} \times \prod_{k=1}^{K} \exp\left\{-\lambda_2 \sum_{i\in C_k} \mathcal{D}(u_i, \mathbf{u}_k)\right\},$$

  such that $\mathbf{C} : |\mathbf{C}| = K$.

- Employ the the scheme in Holmes and Walker (2017) for the selection (λ_1, λ_2).

Sampling

- **Sampling the Fairlets/ Sampling from $[\mathbf{u} \mid \cdot]$.** Efficient scheme utilizing discrete optimal transport and weighted rectangular loop updates.
- **Sampling Clustering Indices/ Sampling from $[\mathbf{C} \mid \cdot]$.**
  - Denote $\mathbf{c}_{-i} = (c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_n)$ is set of clustering indices without the i-th unit.
  - Sample from $\mathbf{P}(c_i = k \mid \mathbf{c}_{-i}, \lambda_1, \mathbf{u})$ via Metropolis updates.

Experiment (Well-specified Case)

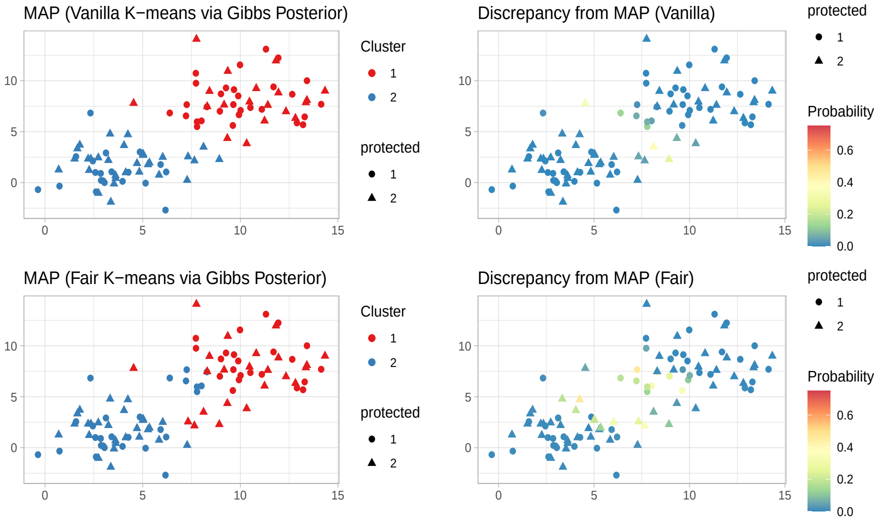


Figure: Set up is same as earlier (multivariate normal components) with $K = 2$.

Experiment (Mis-specified Case)

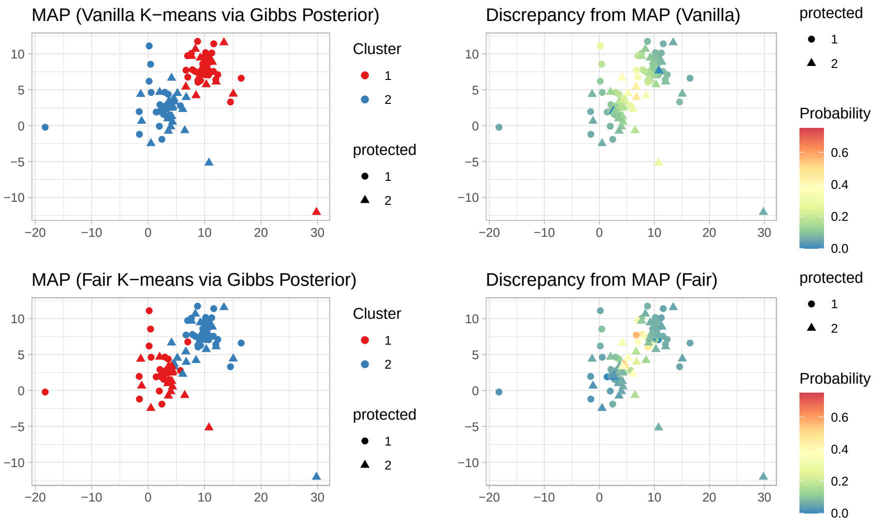
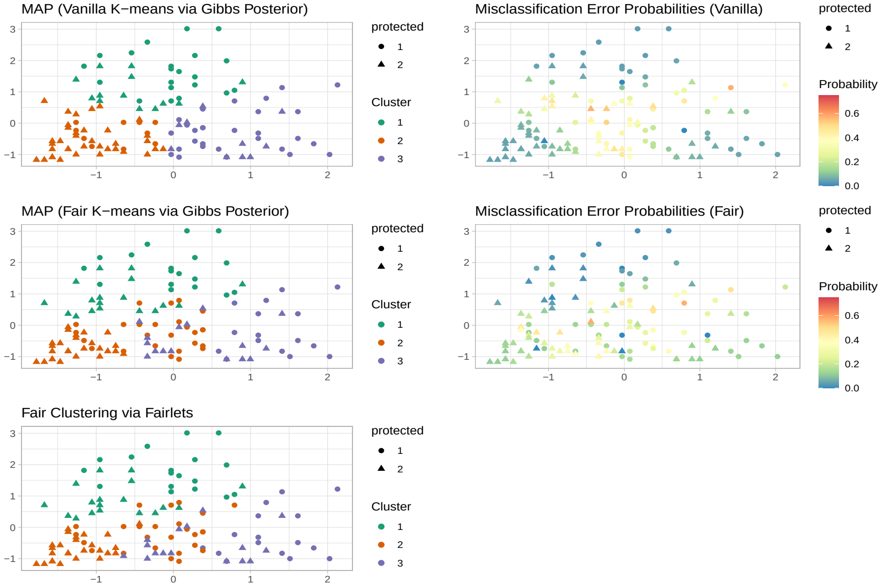


Figure: Set up is same as earlier (multivariate t components) with $K = 2$.

A Benchmark Data (Credit Card Data)



A Look Back

- **Existing Literature.** Uncertainty quantification was largely illusive .
- **HFDP.** Proposed a model-based approach. Developed a concrete notion of optimal recovery and principled performance evaluation.
- **Limitations of HFDP.** Model-based fair clustering frameworks show brittleness under model-misspecification and can be computationally prohibitive.
- **Generalised Bayesian Fair Clustering.** Proposed a framework that is more immune to model-misspecification and support efficient computation.

Check out the Papers!

Figure: **Fair Clustering via Hierarchical Fair-Dirichlet Process.**

Figure: **A Gibbs Posterior Framework for Fair Clustering.**

References I

Ahmadian, S., Epasto, A., Knittel, M., Kumar, R., Mahdian, M., Moseley, B., Pham, P., Vassilvitskii, S., and Wang, Y. (2020a). Fair hierarchical clustering. *CoRR*, abs/2006.10221.

Ahmadian, S., Epasto, A., Kumar, R., and Mahdian, M. (2020b). Fair correlation clustering. *CoRR*, abs/2002.02274.

Bandyapadhyay, S., Inamdar, T., Pai, S., and Varadarajan, K. R. (2019). A constant approximation for colorful k-center. *CoRR*, abs/1907.08906.

Bissiri, P. G., Holmes, C. C., and Walker, S. G. (2016). A general framework for updating belief distributions. *J R Stat Soc Series B Stat Methodol*, 78(5):1103–1130. Epub 2016 Feb 23.

BÃ¶hm, M., Fazzone, A., Leonardi, S., and Schwiegelshohn, C. (2020). Fair clustering with multiple colors.

Chakrabarty, D. and Negahbani, M. (2021). Better algorithms for individually fair k-clustering. *CoRR*, abs/2106.12150.

Chen, X., Fain, B., Lyu, C., and Munagala, K. (2019). Proportionally fair clustering. *CoRR*, abs/1905.03674.

References II

Chierichetti, F., Kumar, R., Lattanzi, S., and Vassilvitskii, S. (2017). Fair clustering through fairlets. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Dahl, D. (2006). Model-based clustering for expression data via a dirichlet process mixture model, in bayesian inference for gene expression and proteomics. *Cambridge University Press*.

Esmaeili, S., Brubach, B., Tsepenekas, L., and Dickerson, J. (2020). Probabilistic fair clustering. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12743–12755. Curran Associates, Inc.

Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., and Venkatasubramanian, S. (2015). Certifying and removing disparate impact. In *KDD*, pages 259–268.

Holmes, C. C. and Walker, S. G. (2017). Assigning a value to a power likelihood in a general bayesian model. *Biometrika*, 104(2):497–503.

Kleindessner, M., Awasthi, P., and Morgenstern, J. (2020). A notion of individual fairness for clustering.

Kleindessner, M., Samadi, S., Awasthi, P., and Morgenstern, J. (2019). Guarantees for spectral clustering with fairness constraints.

References III

Mahabadi, S. and Vakilian, A. (2020). Individual fairness for k-clustering.

Miller, J. W. and Harrison, M. T. (2013). Exact sampling and counting for fixed-margin matrices. *The Annals of Statistics*, 41(3):1569 – 1592.

Rösner, C. and Schmidt, M. (2018). Privacy preserving clustering with constraints. *CoRR*, abs/1802.02497.

Wang, G. (2020). A fast MCMC algorithm for the uniform sampling of binary matrices with fixed margins. *Electronic Journal of Statistics*, 14(1):1690 – 1706.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.