Extending Chamberlain's Law for AI Fairness: A Probabilistic Framework for Quantifying Structural Disparities in Algorithmic Decision-Making

Imoleayomide Ajayi

Contemporary AI fairness auditing relies predominantly on post-hoc dispar ity metrics evaluated at the marginal level of a single protected attribute. Such approaches are analytically insufficient when protected groups overlap, when struc tural disadvantage is intersectional, and when observed disparities may be partially explained by legitimate predictive features. This paper formalises Chamberlain's Law—the principle that fairness auditing should proceed by asking, "if all things are equal, what are the numbers we expect to see?"—as a rigorous probabilistic framework for algorithmic fairness assessment. The framework integrates five method ological components: (i) an expected-outcome fairness baseline derived exclusively from legitimate predictors, operationalised via frequentist logistic regression; (ii) a structural fairness model that augments the baseline with protected-group indicators and intersectional interaction terms; (iii) a Bayesian logistic regression model provid ing full posterior uncertainty quantification over group effects and fairness gaps; (iv) application of the inclusion–exclusion principle for computing union-level expected counts across overlapping protected groups, thereby avoiding double-counting errors common in subgroup audits; and (v) a novel Inequality Attribution Score (IAS), defined as the variance share of the linear predictor attributable to identity-related structural factors, providing a single interpretable index of structural inequity. Using five simulated scenarios spanning no bias, marginal bias, intersectional bias, proxy feature bias, and error-rate bias, we demonstrate that marginal fairness metrics can mask substantial intersectional disparities, that the IAS captures structural inequity invisible to conventional metrics, and that Bayesian inference provides practically important uncertainty bounds around all disparity estimates. The framework is positioned as a principled, inferential complement to existing algorithmic auditing practice.