

## Regression under demographic parity constraints via unlabeled post-processing

We address the problem of performing regression while ensuring demographic parity, even without access to sensitive attributes during inference. We present a general-purpose post-processing algorithm that, using accurate estimates of the regression function and a sensitive attribute predictor, generates predictions that meet the demographic parity constraint. Our method involves discretization and stochastic minimization of a smooth convex function. It is suitable for online post-processing and multi-class classification tasks only involving unlabeled data for the post-processing. Unlike prior methods, our approach is fully theory-driven. We require precise control over the gradient norm of the convex function, and thus, we rely on more advanced techniques than standard stochastic gradient descent. Our algorithm is backed by finite-sample analysis and post-processing bounds, with experimental results validating our theoretical findings.

Joint work with G. Taturyan, M. Hebiri