OUTCOME INDISTINGUISHABILITY, SCAFFOLDING SETS, AND PAN-CALIBRATION

CYNTHIA DWORK

Prediction algorithms score individuals, or individual instances, assigning to each one a number in the range from 0 to 1. That score is often interpreted as a probability: What are the chances that this loan will be repaid? How likely is this tumor to metastasize? A key question lingers: What is the "probability" of a non-repeatable event? This is the defining problem of AI. Without a satisfactory answer, how can we even specify what we want from an ideal algorithm?

This talk will introduce *outcome indistinguishability* [2], a desideratum with roots in computational complexity theory, and will situate the concept within the landscape of algorithmic fairness.

Outcome indistinguishability generalizes *multi-calibration*, a fairness notion for prediction algorithms that requires simultaneous calibration on a (possibly large) pre-specified collection of subsets of the population [3]. Here, too, a question lingers: what can be done to ensure that all subordinated groups – including those whose members cannot advocate for themselves – are included in the collection?

We will show how to circumvent this problem through the use of a *Scaffolding Set* collection [1], and give some simple conditions under which such a collection can be efficiently constructed. When these conditions are not met, no harm is done; when they are satisfied, calibration is achieved simultaneously on all large subpopulations, a concept we call *pan-calibration*.

References

[1] M. Burhanpurkar, Z. Deng, C. Dwork, and L. Zhang, *Scaffolding Sets*, arXiv preprint arXiv:2111.03135, 2021.

[2] C. Dwork, M. P. Kim, O. Reingold, Guy N. Rothblum, and G. Yona, *Outcome Indistinguishability*, Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 1095–1108, 2021

[3] U. Hébert-Johnson, M. P. Kim, O. Reingold, and G. Rothblum, *Multicalibration: Calibration for the (computationally-identifiable) masses*, International Conference on Machine Learning, PMLR, 1939–1948, 2018

Harvard University
Email address: dwork@seas.harvard.edu