On Fairness and Calibration

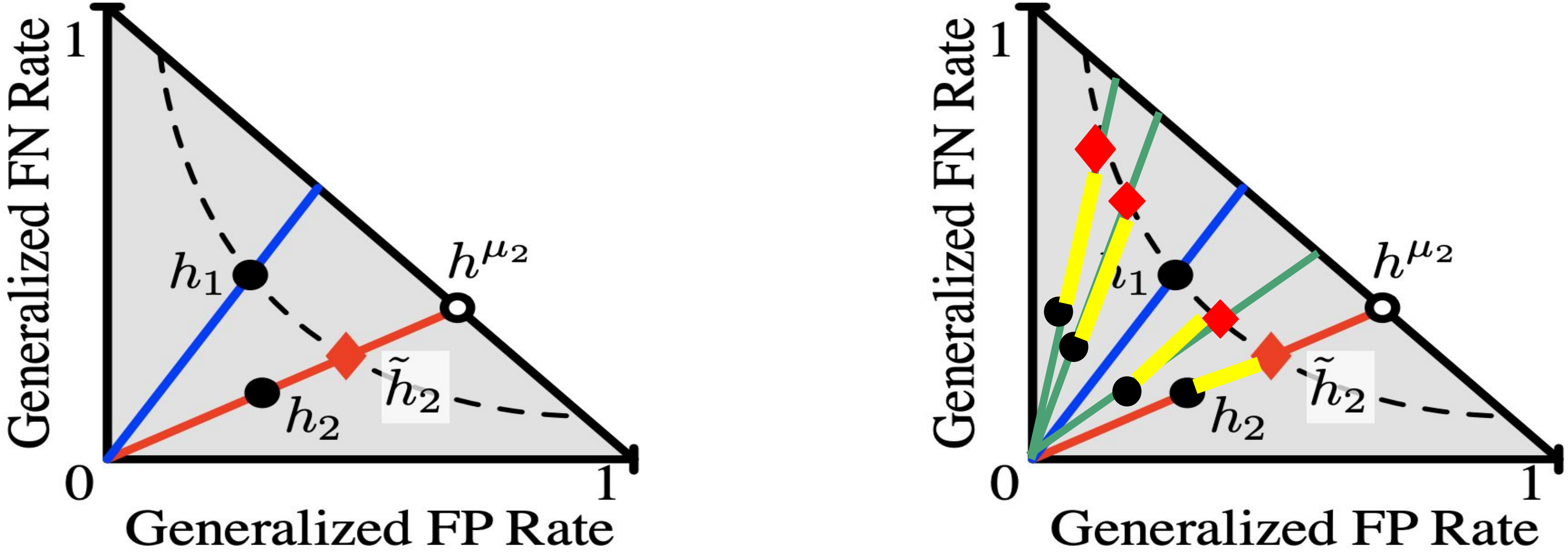
Critique

Aditya Jain, Manish Reddy

Key Takeaways

- (Mostly) Everything is Definitions or Theorems
- Provides an **intuitive geometric** framework to understand **fairness using Calibration and Equalized Odds (or Relaxed Equalized Odds).**
- **Is Fairness feasible ?** Infeasible when the best classifiers are not far from the trivial classifiers (leaving little room for interpolation). In that case, **more data or features** required to achieve their fairness notion!
- Clearly states and proves competing notions of fairness such as
  - Calibration vs EO
  - Group vs Individual Fairness
  - Seeking equality for a single error rate (e.g. false-negatives,) will necessarily increase disparity with respect to the other error.

Practical Concern: Many Groups/ Continuous variable



A single badly classified group leads to increased costs of all other groups and the system as a whole.

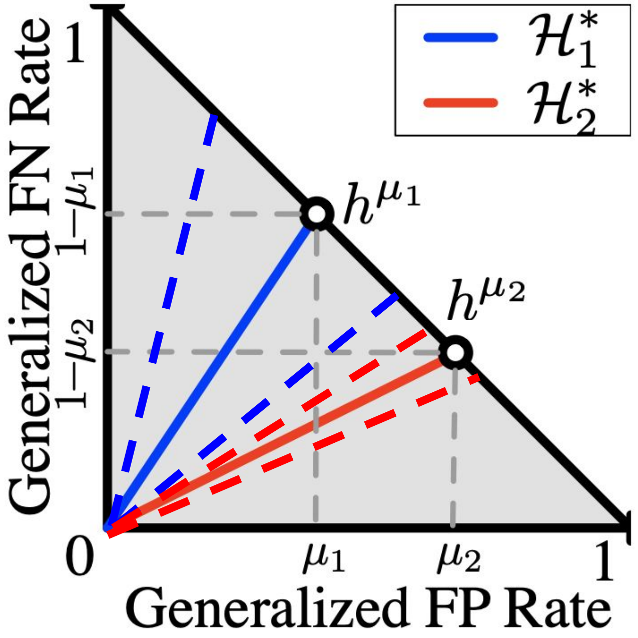
Relaxing Calibration → General Comment

A natural question of relaxing calibration arises which is discussed in the supplementary section!.

Groups have **unequal number of datapoints**

(Relaxation)ε ∝ 1/(Number of datapoints)

Underlying concept : Estimates across different groups have different uncertainty



Practical Concern - 2

- Individual vs Group Tradeoff

$$\tilde{h}_2(\mathbf{x}) = \begin{cases} h^{\mu_2}(\mathbf{x}) = \mu_2 & \text{with probability } \alpha \\ h_2(\mathbf{x}) & \text{with probability } 1 - \alpha \end{cases}$$

- Not equitable
- How would people receive such a alpha-coin toss?

Choice of cost function:

$$g_t(h_t) = a_t c_{fp}(h_t) + b_t c_{fn}(h_t)$$

- Linearity is nice, but can we have more powerful expressions with non-linearity? (Like MAX and MIN)
- How would the Algorithms + Convexity change for other cost-functions?

Nitpicking!

- Experimental Settings
  - Experiment 2: Health prediction - Hard coded r_fp and r_fn.
    - Reasoning?
    - Alternatives?