Bias and Fairness - A framework proposal

Jakob Kisiala
True North Partners LLP | tnp.eu
jakob.kisiala@tnp.eu

Introduction

At TNP, we looked at different ideas about bias and fairness measurement and enhanced them to arrive at a framework proposal that can be used by financial institutions to reduce bias and improve fairness in lending decisions. We provide an illustrative example with a public dataset and reasonable assumptions.

Bias vs Fairness

TNP propose to adopt a definition to bias that is an objectively observable state, while fairness includes a judgement as to what is perceived as "right":

*Algorithm **bias** refers to systematic decisions in systems that lead to skewed outcomes*	***Fairness** is to treat people equally, or in a way that is right and reasonable, and to make decisions that avoid unjustified adverse effects*

Individual vs group level measures

Bias and fairness can be measured for individuals or groups. In the context of credit decisioning for a retail portfolio, we are focussing on group level measures here:

Statistical Parity ($\hat{Y} \perp D$) All customers should get the same decision, regardless of their group membership.	**Conditional Statistical Parity** ($\hat{Y} \perp D \vert X_c$) All customers should get the same decision, when accounting for another variable (e.g. income).	**Equalised Odds** ($\hat{Y} \perp D \vert Y$) Decisions should be the same across groups, when splitting by good / bad outcomes.
Equal Opportunity ($\hat{Y} \perp D \vert Y = 1$) All creditworthy customers should get the same decision, regardless of group membership.	**Predictive parity** ($Y \perp D \vert \hat{Y} = 1$) All accepted customers should perform similarly, regardless of group membership.	**Accuracy Parity** ($\hat{Y} = Y \perp D$) Predictions should align with actual outcomes, regardless of group membership.

Fairness appetite statement

Similar to risk appetite statements, a fairness statement should be formulated, based on the previous measures. A possible example is:

- We want to have **equalised odds** for our credit decisions to be within 80% *of each other* for men and women.
- In other words, the TPR and FPR for women should at least be 80% of the TPR and FPR for men.

Fairness budget

In situations where the aim is to counter historic, systemic economic differences (e.g. pay gaps affecting repayment capacity) a fairness budget might need to be put aside.

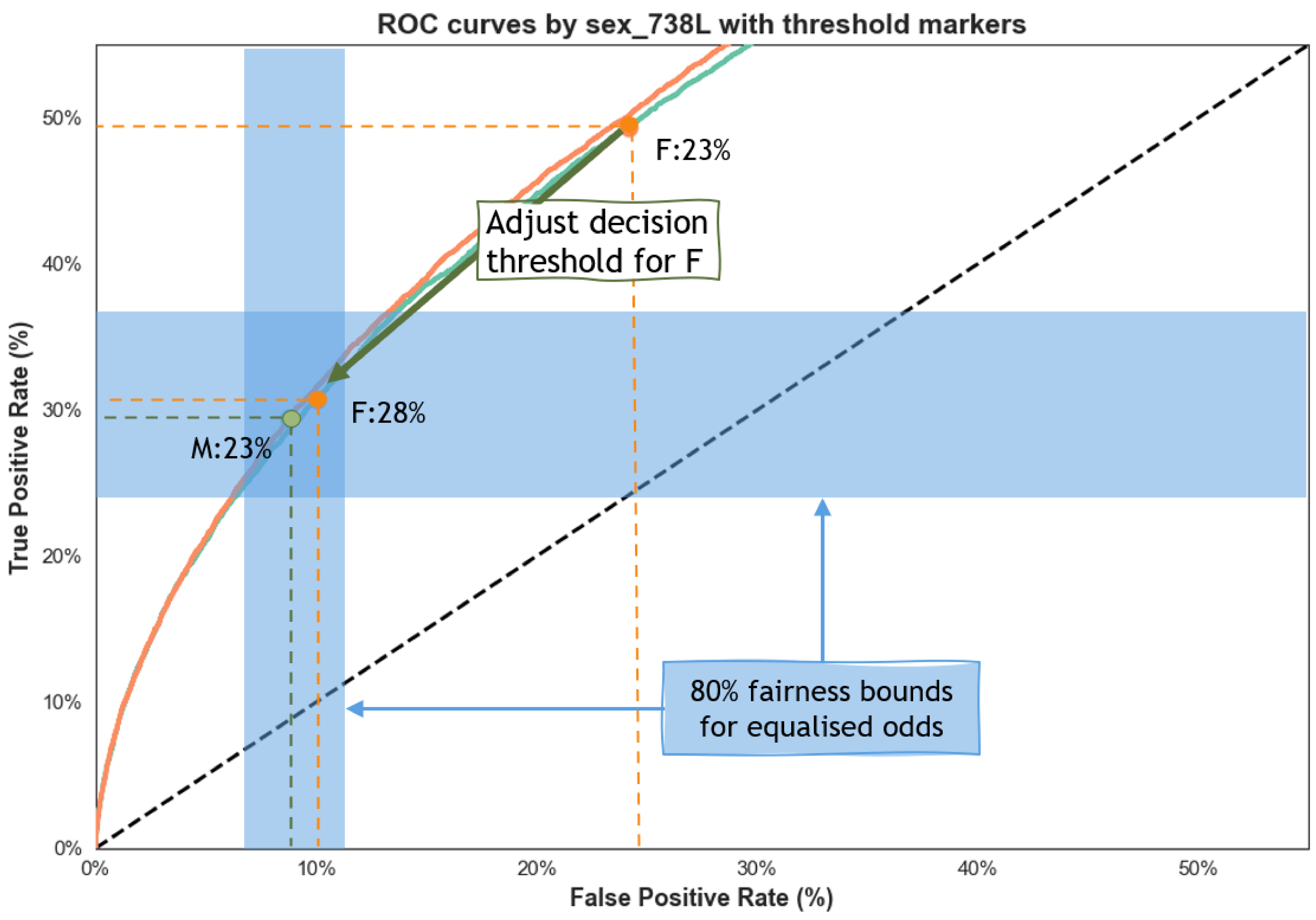
Strategies to achieve fairness

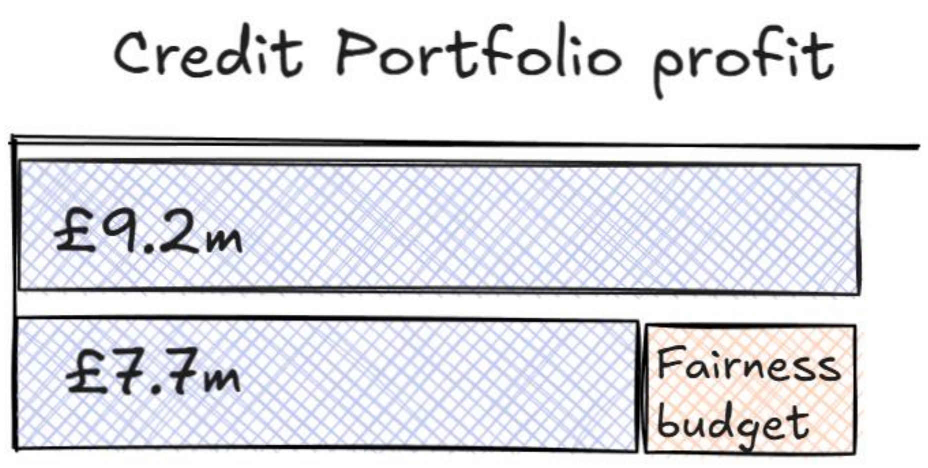
To reach the desired fairness appetite with the fairness budget, 4 possible strategies can be applied:

Pre-processing (adjust input data) Resample input data or remove correlated features	**In-processing (adjust model fitting process)** Penalise missed fairness metric in addition to modelling error
Post-processing (adjust model outputs) Change model parameters after training is completed	**No processing (adjust decision thresholds)** Set different decision thresholds for different groups

*Example impacts**

- Aiming to achieve equalised odds within 80% for men and women,
- by adjusting the credit decision cut off,
- requires a fairness budget of ~ £1.5m.





* using a sample dataset, with a biased baseline model and several assumptions around credit decision strategy and P/L drivers

TRUE NORTH PARTNERS
FINANCE | RISK | STRATEGY