Equalized conditional error rates vs marginal calibration

Daniel Hsu

November 12, 2025

Consider jointly-distributed $\{0,1\}$-valued random variables $(A, Y, \hat{Y})$. Here, $\hat{Y}$ is regarded as a prediction of Y. For each $a \in \{0,1\}$, define

- *false positive rate for group a*:

$$\mathrm{FPR}_a := \Pr(\hat{Y} = 1 \mid Y = 0, A = a);$$

- *false negative rate for group a*:

$$\mathrm{FNR}_a := \Pr(\hat{Y} = 0 \mid Y = 1, A = a);$$

- *base positive rate for group a*:

$$\mathrm{BPR}_a := \Pr(Y = 1 \mid A = a);$$

and we say $\hat{Y}$ is *marginally calibrated for group a* if

$$\Pr(\hat{Y} = 1 \mid A = a) = \mathrm{BPR}_a\,.$$

Theorem 1 (Chouldechova; Kleinberg-Mullainathan-Raghavan)**.** *Unless*

$$\mathrm{BPR}_0 = \mathrm{BPR}_1 \quad \textit{or} \quad \mathrm{FPR}_0 = \mathrm{FPR}_1 = \mathrm{FNR}_0 = \mathrm{FNR}_1 = 0,$$

it is impossible for all of the following to simultaneously hold:

1. $\mathrm{FPR}_0 = \mathrm{FPR}_1$*;*

2. $\mathrm{FNR}_0 = \mathrm{FNR}_1$*;*

3. $\hat{Y}$ *is marginally calibrated for group* a*, for each* $a \in \{0,1\}$*.*

Proof. Assume that $\mathrm{FPR}_0 = \mathrm{FPR}_1$, $\mathrm{FNR}_0 = \mathrm{FNR}_1$, and that $\hat{Y}$ is marginally calibrated for each group $a \in \{0,1\}$. We need to show that either the groups have the same base positive rates, or the false positive rates and false negative rates for each group are zero.

For each $a \in \{0,1\}$, the following chain of equalities are implied by the assumptions:

$$\begin{aligned}
\mathrm{BPR}_a &= \Pr(\hat{Y} = 1 \mid A = a) \quad \text{(since } \hat{Y} \text{ is marginally calibrated for group } a\text{)} \\
&= \Pr(\hat{Y} = 1 \mid Y = 0, A = a) \cdot \Pr(Y = 0 \mid A = a) \\
&\qquad + \Pr(\hat{Y} = 1 \mid Y = 1, A = a) \cdot \Pr(Y = 1 \mid A = a) \\
&= \mathrm{FPR}_a \cdot (1 - \mathrm{BPR}_a) + (1 - \mathrm{FNR}_a) \cdot \mathrm{BPR}_a \\
&= \mathrm{FPR} \cdot (1 - \mathrm{BPR}_a) + (1 - \mathrm{FNR}) \cdot \mathrm{BPR}_a \quad \text{(since } \mathrm{FPR}_0 = \mathrm{FPR}_1 \text{ and } \mathrm{FNR}_0 = \mathrm{FNR}_1\text{)},
\end{aligned}$$

where FPR denotes the common value of FPR_0 and FPR_1, and FNR denotes the common value of FNR_0 and FNR_1.[1] Simplifying the result above, we obtain the following system of equations:

$$\begin{aligned}
\mathrm{FPR} \cdot (1 - \mathrm{BPR}_0) &= \mathrm{FNR} \cdot \mathrm{BPR}_0 \\
\mathrm{FPR} \cdot (1 - \mathrm{BPR}_1) &= \mathrm{FNR} \cdot \mathrm{BPR}_1 \,.
\end{aligned}$$

Together, these equations imply

$$(\mathrm{FPR} + \mathrm{FNR}) \cdot (\mathrm{BPR}_0 - \mathrm{BPR}_1) = 0.$$

This means that either $\mathrm{BPR}_0 = \mathrm{BPR}_1$ or $\mathrm{FPR} + \mathrm{FNR} = 0$. $\square$

[1] Note that because $\mathrm{FPR}_0 = \mathrm{FPR}_1 = \mathrm{FPR}$ and $\mathrm{FNR}_0 = \mathrm{FNR}_1 = \mathrm{FNR}$, we have

$$\begin{aligned}
\Pr(\hat{Y} = 1 \mid Y = 0) &= \mathrm{FPR}_0 \cdot \Pr(A = 0 \mid Y = 0) + \mathrm{FPR}_1 \cdot \Pr(A = 1 \mid Y = 0) \\
&= \mathrm{FPR} \cdot \Pr(A = 0 \mid Y = 0) + \mathrm{FPR} \cdot \Pr(A = 1 \mid Y = 0) \\
&= \mathrm{FPR},
\end{aligned}$$

and similarly, $\Pr(\hat{Y} = 0 \mid Y = 1) = \mathrm{FNR}$. That is, FPR and FNR are, respectively, the false positive rate and false negative rate for the overall population.