1 Standard implementation

Following each classification $C \in \{\text{LENS}, \text{NOT}\}$, the original SWAP algorithm derives the posterior probability that a subject contains a gravitational lens using a simple Bayesian update rule.

$$\Pr(\text{LENS}|C) = \frac{\Pr(C|\text{LENS}) \cdot \Pr(\text{LENS})}{[\Pr(C|\text{LENS}) \cdot \Pr(\text{LENS}) + \Pr(C|\text{NOT}) \cdot \Pr(\text{NOT})]} \quad (1)$$

Where $\Pr(C|\text{LENS})$ and $\Pr(C|\text{NOT})$ are the relevant elements of the classifiers *confusion matrix*.

Naïve application of (1) permits the undesirable situation whereby a nominally *perfect* classifier (for whom $\Pr(C|\text{LENS}) = 1$ and $\Pr(C|\text{NOT}) = 0$) may immediately promote the posterior probability of a subject to unity with a single LENS classification. Similarly, a single NOT classification by this classifier would immediately demote the posterior probability to zero.

Either situation would result in immediate retirement of the subject from the classification set with no possiblity of reinstatement.

2 Alternative implementations

2.1 Softening the response to perfect classifiers

A simple strategy that would avoid the pathology oulined in §1 entails modification of (1) by simple multiplicative scaling of $\Pr(C|\text{LENS})$ and $\Pr(C|\text{NOT})$ such that:

$$\Pr(\text{LENS}|C) = \quad (2)$$
$$\frac{\alpha \cdot \Pr(C|\text{LENS}) \cdot \Pr(\text{LENS})}{[\alpha \cdot \Pr(C|\text{LENS}) \cdot \Pr(\text{LENS}) + \alpha \cdot \Pr(C|\text{NOT}) \cdot \Pr(\text{NOT})]}$$

where $\alpha < 1$ is a suitably chosen scalar value.

There are obvious shortcomings for this extremely simple approach, including (but likely not limited to)

1. If α is excessively small, the number of required number of matching classifications for a subject to pass either of its retirement thresholds may become excessively large.

2. Penalizing affirmative (LENS) and negative (NOT) by the same factor may not be appropriate. It may be far easier to affirm the presence of a lens than it is to exclude it.

3. The experience of the classifier, quantified by the **number** (N_C) of subjects that they have classified is not **explicitly** accounted for. In the original framework a classifier may obtain a *perfect* confusion matrix after classifying a **single** gold-standard subject.

4. The difficulty of classifying a *particular* subject is not accounted for. If additional *metadata* (Θ), describing properties of the subject image are available, it may be sensible to penalize or promote a classification's weight appropriately.

Considering these shortcomings alone, a sensible reformulation of (2) might entail replacement of a single scalar-valued α with a pair of scalar-valued *functions* $(\alpha_{\text{LENS}}(N_C, \Theta), \beta_{\text{NOT}}(N_C, \Theta)$, such that

$$\Pr(\text{LENS}|C, N_C, \Theta) = \tag{3}$$
$$\frac{\alpha_{\text{LENS}}(N_C, \Theta) \cdot \Pr(C|\text{LENS}) \cdot \Pr(\text{LENS})}{[\alpha_{\text{LENS}}(N_C, \Theta) \cdot \Pr(C|\text{LENS}) \cdot \Pr(\text{LENS}) + \beta_{\text{NOT}}(N_C, \Theta) \cdot \Pr(C|\text{NOT}) \cdot \Pr(\text{NOT})]}$$

While the functional forms of α_{LENS} and β_{NOT} are arbitrary, the requirement that $0 \geq \Pr(\text{LENS}|C, N_C, \Theta) \leq 1$ mandates that $\alpha_{\text{LENS}} < 1$ and $\beta_{\text{NOT}} < 1 \; \forall \; N_C, \Theta$.