Effects of the imbalance in the Log LR Cost Function (C_{llr})

Arts & Humanities Research Council

voice and identity 0101101ID1

Vincent Hughes

vincent.hughes@york.ac.uk

THE UNIVERSITY of York

1. Introduction

Essential elements of the *paradigm shift* across forensic science (Morrison 2014):

(i) use of the **LR**

(ii) testing **validity** (accuracy) & **reliability** (precision): conducted using sets of SS and DS pairs where:

- N SS pairs = N speakers
- N DS pairs = min. (N speakers^2)-N speakers

Validity metrics:

Equal Error Rate (EER):

- false hits (DS LLR>0) == misses (SS LLR<0)
- *accept-reject* decisions based on $p(H|E)$

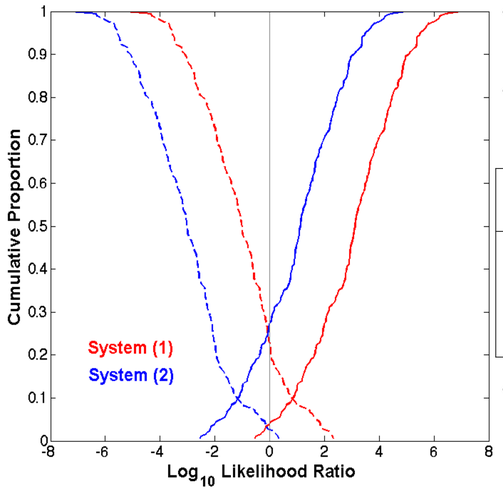
Log LR Cost Function (C_{llr}):

$$C_{llr} = \frac{1}{2}\left(\frac{1}{N_{ss}}\sum_{i=1}^{N_{ss}}\log_2\left(1+\frac{1}{LR_{ss_i}}\right)+\frac{1}{N_{ds}}\sum_{i=1}^{N_{ds}}\log_2\left(1+LR_{ds_i}\right)\right)$$

- based on magnitude of *contrary-to-fact* LLRs
- mean of two cost functions calculated for SS & DS pairs independently

How does the imbalance in N SS and N DS pairs affect C_{llr}?

Which system is best?



- 2 systems (20 SS/ 180 DS)
- differing only in % false hits/ misses

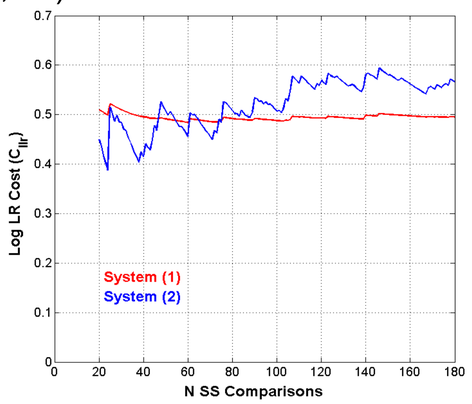
	(1)	(2)
EER	10%	10%
C_{llr}	0.50	0.44

- C_{llr} lower for system with higher % misses

3. Experiments

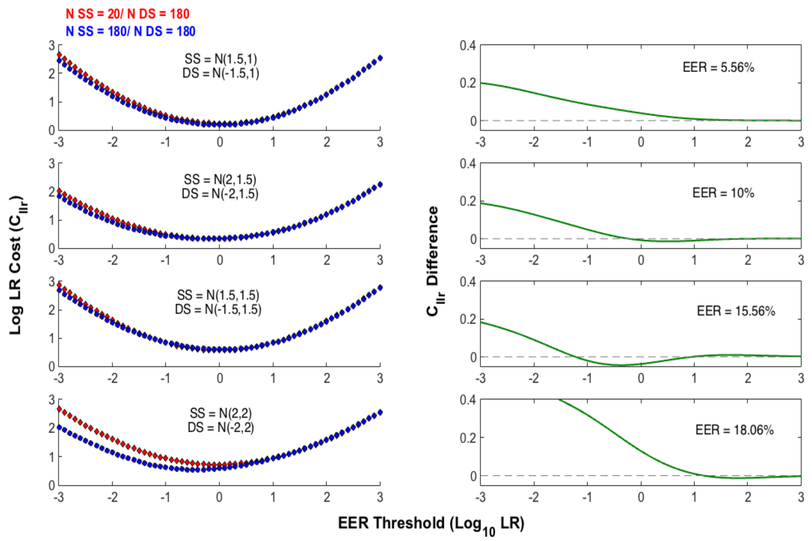
3.1 Experiment 1: what if N SS == N DS?

- hypothetical LLRs from normal distributions: SS = $N(2,1.5)$ & DS = $N(-2,1.5)$
- **system (1): LLR+1**
  - N SS (20:180) == N DS (20:180)
- **system (2): LLR-1**
  - N SS (20:180) ≠ N DS (180)
- C_{llr} calculated as N SS increased from 20 to 180



3.2 Experiment 2: how does the threshold for EER affect C_{llr} when N SS and N DS are (im)balanced?

- hypothetical LLRs from normal distributions with four EERs (5.56%, 10%, 15.56%, 18.06%)
- from each EER condition, two systems created:
  - N SS (180) == N DS (180)
  - N SS (20) ≠ N DS (180)
- C_{llr} calculated as EER threshold shifted from -3 to +3



4. Discussion

- systems with higher % misses generate better C_{llr} than systems with higher % false hits
  - N SS LLRs < N DS LLRs ∴ cost is lower

Experiment 1:

- smaller N SS (as is the case in system testing) generates better C_{llr} than if N SS == N DS
- absolute N pairs doesn't affect C_{llr} if N SS == N DS (as long as there are a sufficient number of comparisons to calculate validity in the first place)

Experiment 2:

- as in Exp 1, systems with imbalanced N SS and N DS generate better C_{llr}
- magnitude of C_{llr} differences between balanced and imbalanced systems dependent on:
  - **EER:** systems with higher EER will show a greater effect for SS/DS imbalance)
  - **LLR value at the EER:** the further the EER threshold from 0 (towards negative values), the greater the effect of SS/DS imbalance

What use is validity anyway?

- validity = means of comparing multiple systems (e.g. different input variables, modelling techniques…)
- **but** the absolute validity of the system used in a case is not of use to the trier-of-fact (in fact, it may be misleading)
  - **validity is a property of the system rather than the LR for a single comparison**
  - more useful for the court to have a credible interval (CI) or the LR of the LR

5. Conclusion

- when comparing systems it is worth considering the performance when N SS == DS, rather than just the absolute C_{llr}