Deriving Confidence Intervals of Log-likelihood Ratio

Nik Suki

Department of Computing

University of Surrey

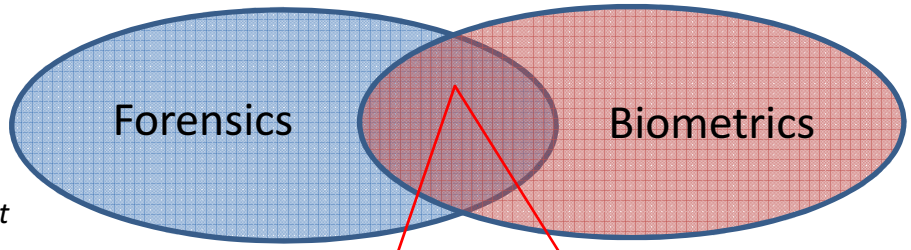
Outline

- Introduction
  - Motivation
- The proposed algorithm
  - Bootstrapping
  - Logistic Regression
- Experimental results
- Summary and future work

Introduction

- Forensic Science and Biometrics

The use of science (applications or technical methods) to the investigation or establishment of facts or evidence in the court of law.



The science of establishing identity of individuals based on their biological and behavioural characteristics.

The usage of human based characteristics to analyse and interpret facts or evidence for the purpose of law.

- Biometric system scores = Forensic Evidence (E) [1]

1. D. Meuwly and R. Veldhuis, "Forensic biometrics: From two communities to one discipline," Proceedings of the BIOSIG 2012, International Conference of the Biometrics Special Interest Group- (BIOSIG) on, 2012.

- Scientific assessment of the performance of any processes involved in forensic science
  - Critical, increasing importance since ***Daubert rules***
- Performance of evidence evaluation methods should be measured [4]:
  - Value of the evidence: Likelihood Ratio

Whether the theory can be **tested** and has been tested.

Whether the technique has been published or subjected to **peer** review.

Whether the technique has a known or potential rate of **error** in application.

Whether **standards** exist and are maintained to control the operation of the technique.

Whether the technique is generally accepted within the relevant scientific community (Frye)

Whether the technique is based on facts or data of a type reasonably relied on by experts in the field.

Whether the technique has a probative value that is not outweighed by the dangers of unfair prejudice, confusion of issues or misleading the jury.

4. Daniel Ramos and Joaquin Gonzalez-Rodriguez, "Reliable support: Measuring calibration of likelihood ratios," Forensic Science International, vol. 230, no. 13, pp. 156 – 169, 2013, fEAFSg 2012 6th European Academy of Forensic Science Conference The Hague, 20-24 August 2012.

- Bayesian Framework for Interpretation of Evidence (E):

The "combined" opinion

The contribution of forensic investigator

The opinion of the court before the evidence is presented

$$\frac{p(H_0|E)}{p(H_1|E)} = \frac{p(E|H_0)}{p(E|H_1)} \cdot \frac{p(H_0)}{p(H_1)}$$

Posterior odds | Likelihood Ratio | Prior odds

Evidential value of E

- Evidential value of E, Log-likelihood Ratio of evidence (LLR(E)) in log domain;

$$\boldsymbol{LLR(E) = log \frac{p(E|H_0)}{p(E|H_1)}}$$

The strength of evidence (E)

H_0, prosecution hypothesis: The suspect *was* ***present*** at the crime scene based on E

H_1, defence hypothesis: The suspect *was* ***not present*** at the crime scene based on E

Motivation

- Problem: ***Misleading Evidence*** [3]
- Support wrong proposition

  e.g., lack of the database used as populations, mismatch in conditions of elements in population database and in the evidence and degraded quality or quantity of the evidential materials [4]-[7]

  Worst case scenario → prosecute an innocent people

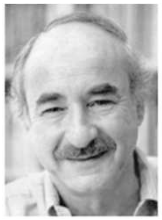
- Performance of evidence evaluation methods should be measured:
  - Value of the evidence: Likelihood Ratio
  - Performance of LLR(E)? How and what to measure?

3. Richard Royall, "On the probability of observing misleading statistical evidence," Journal of the American Statistical Association, vol. 95, no. 451, pp. 760–768, 2000.

Proposed Algorithm

- Performance evaluation for LLR(E)

The Statistician 32 (1983)
The Comparison and Evaluation of Forecasters
MORRIS H. DeGROOT and STEPHEN E. FIENBERG




Journal of the American Statistical Association
September 1982, Volume 77, Number 379
The Well-Calibrated Bayesian
A. P. DAWID*

J. R. Statist. Soc. A (1979), 142, Prt 2, pp. 146-180
On the Reconciliation of Probability Assessments
D. V. LINDLEY, A. TVERSKY and R. V. BROWN



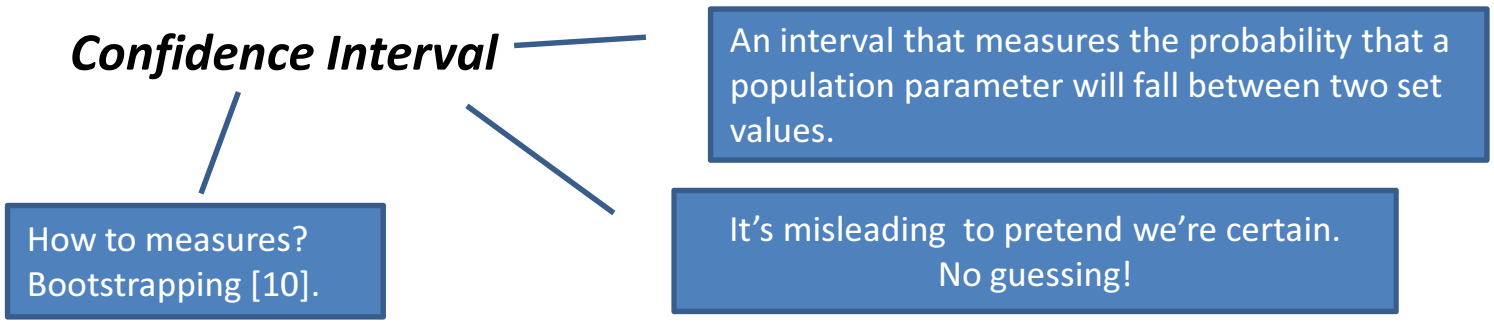
Forensic Science International, vol. 230, no. 13, pp. 156-169, 2013
Reliable support: Measuring calibration of likelihood ratios
D. RAMOS and J. G. RODRIGUEZ

- ***Calibration*** is described as desired properties.

How accurate the event will occur…

Objectives of experiment

- What is the behaviour of confidence interval of LLR(E) for all different E values?
  1. Due to difference bootstrapping methods
  2. And difference approaches to calculate LLR(E)
     - *Density Estimation vs Logistic Regression*



10. Norman Poh, Alvin Martin, and Samy Bengio, "Performance generalization in biometric authentication using joint user-specific and sample bootstraps," IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), July 2005.

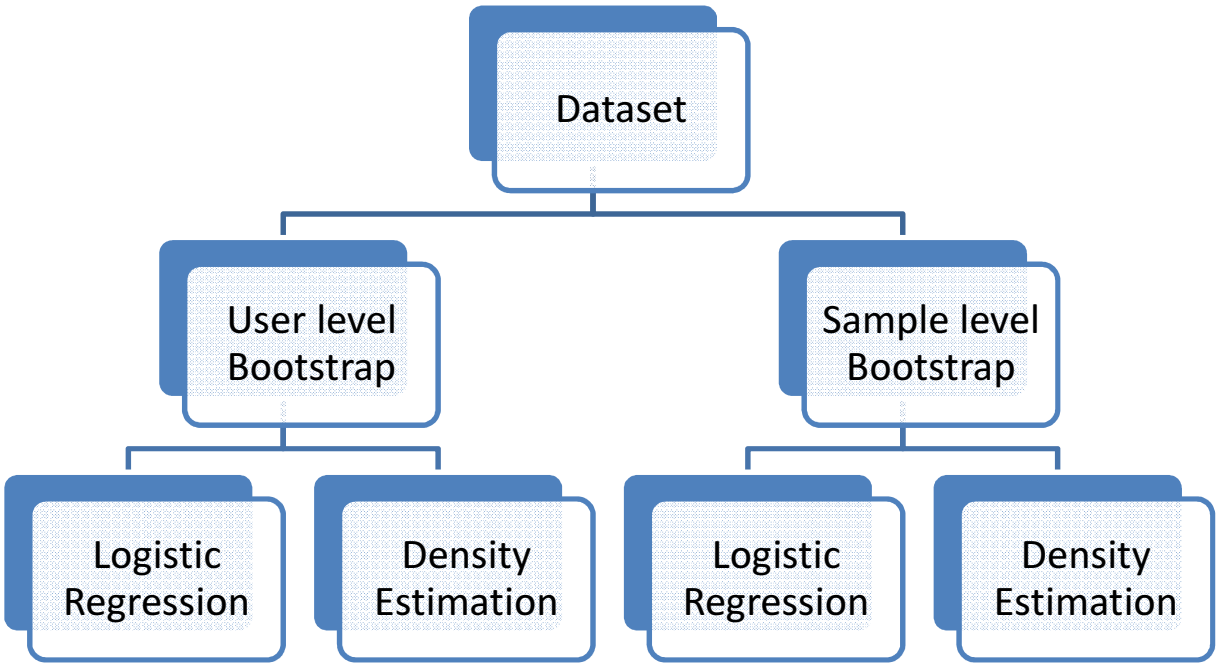
LLR(E)

- Ways to calculate LLR(E)
  - Logistic Regression

    As a calibration method

  - Density Estimation
    - i.e.: Kernel Density Function

Overall Procedure



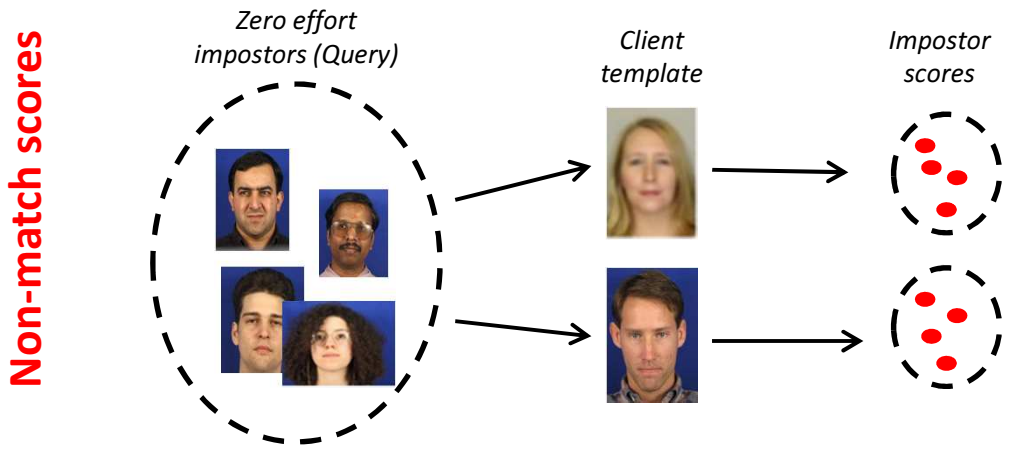
Dataset

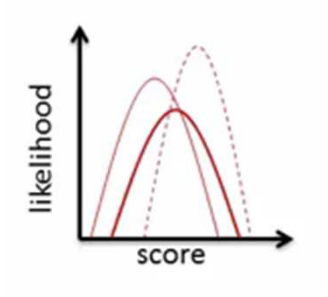
- XM2VTS Database
  - Multimodal biometric database **13**
    - 7 face database
    - 6 speech database
  - 200 users and each user has two **genuine** samples and 600 impostor samples

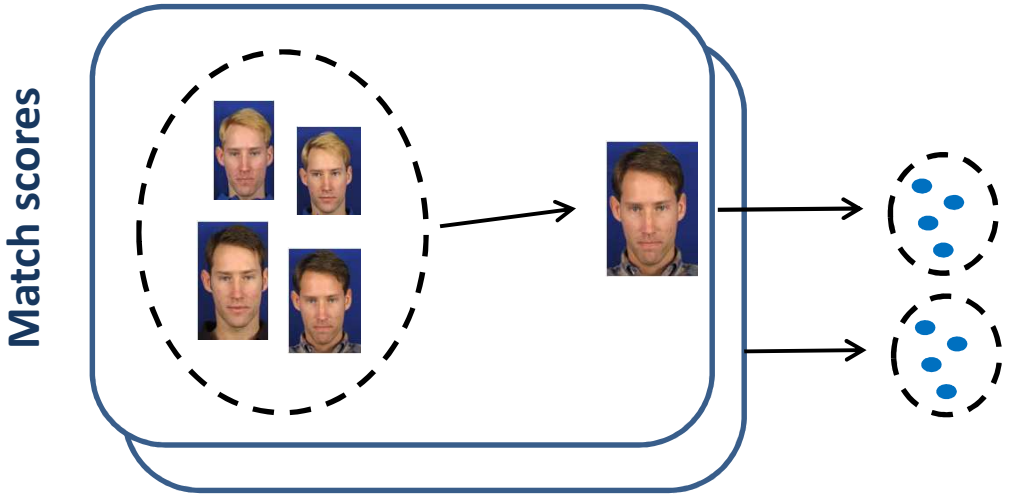
Match scores

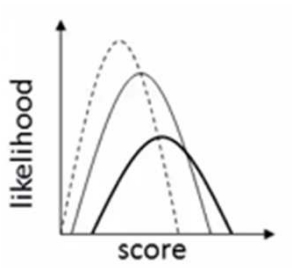
Non-Match scores

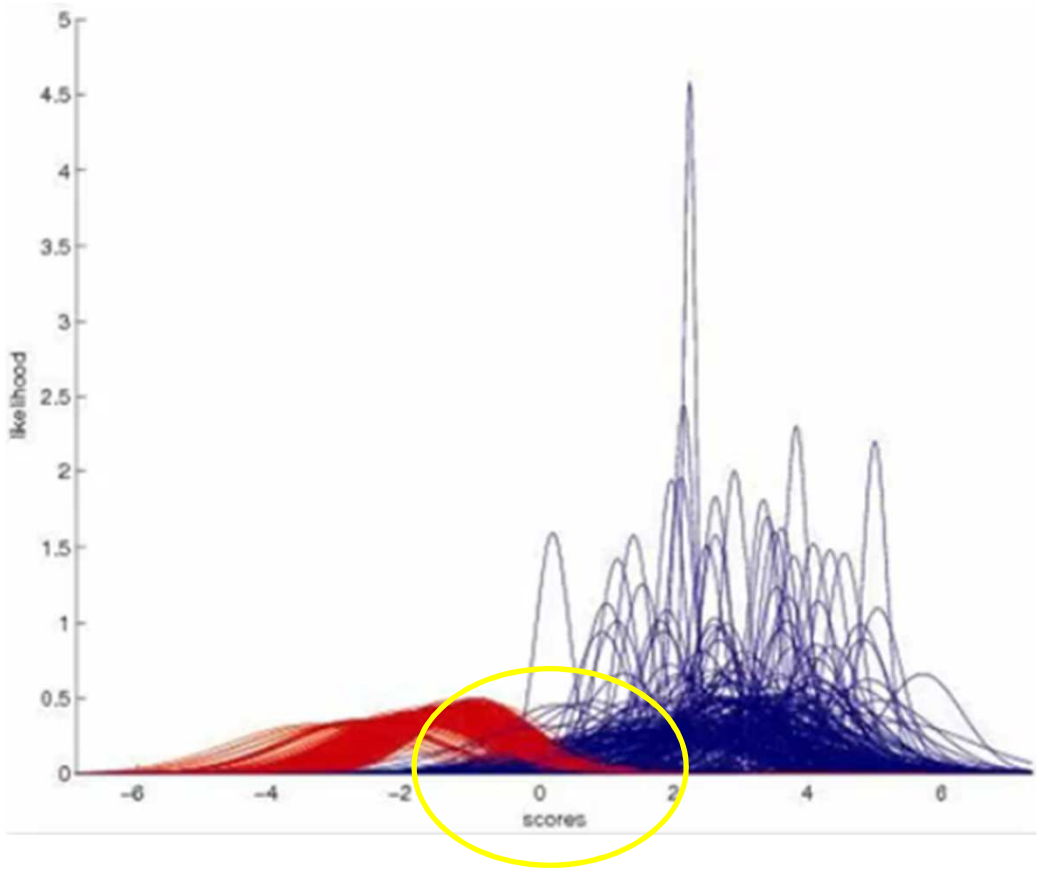
18. J. Luettin and G. Maitre, “Evaluation protocol for the XM2VTS database,” 1998.


Non-match scores
Zero effort impostors (Query)
Client template
Impostor scores

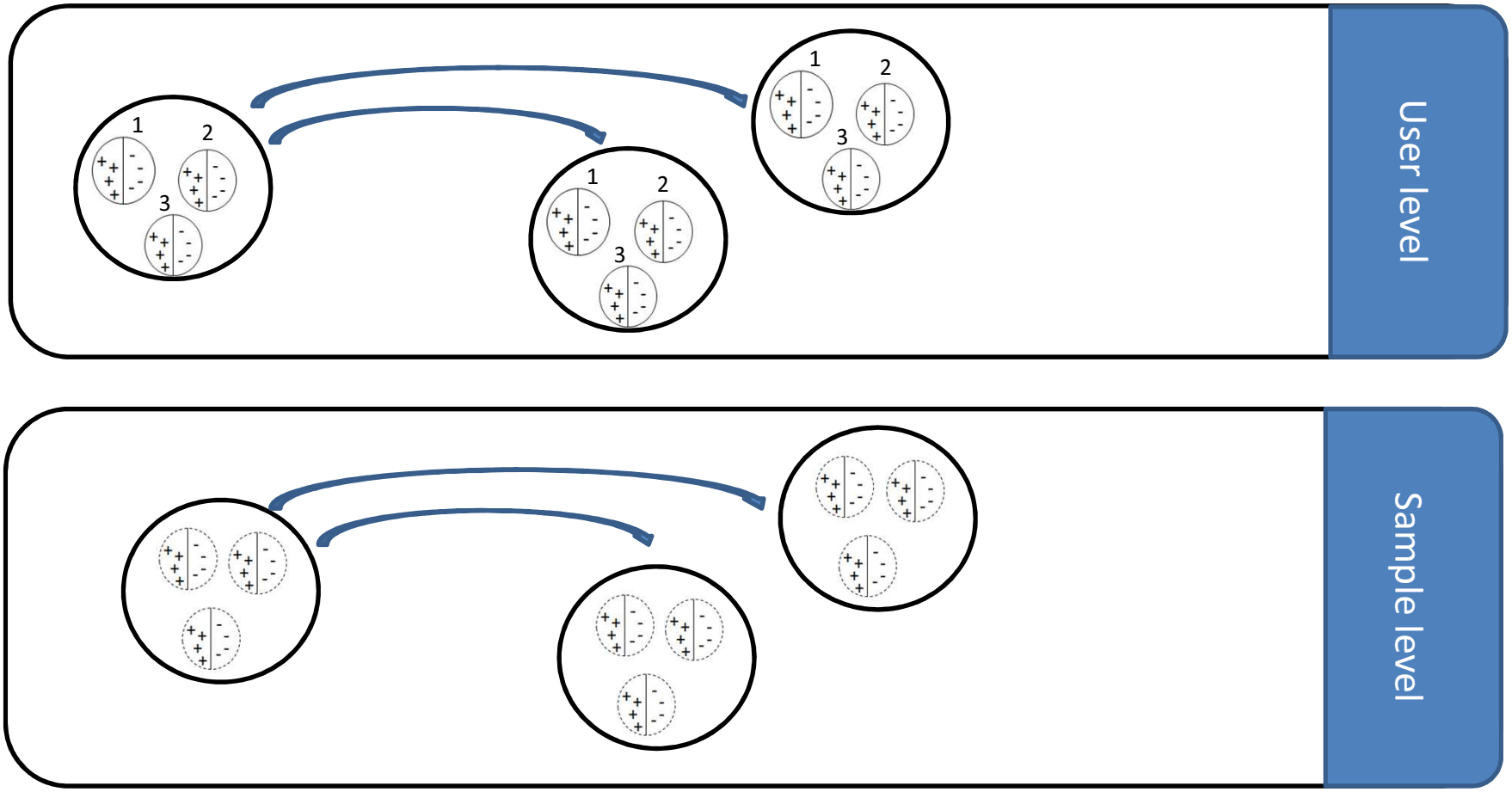

likelihood
score


Match scores

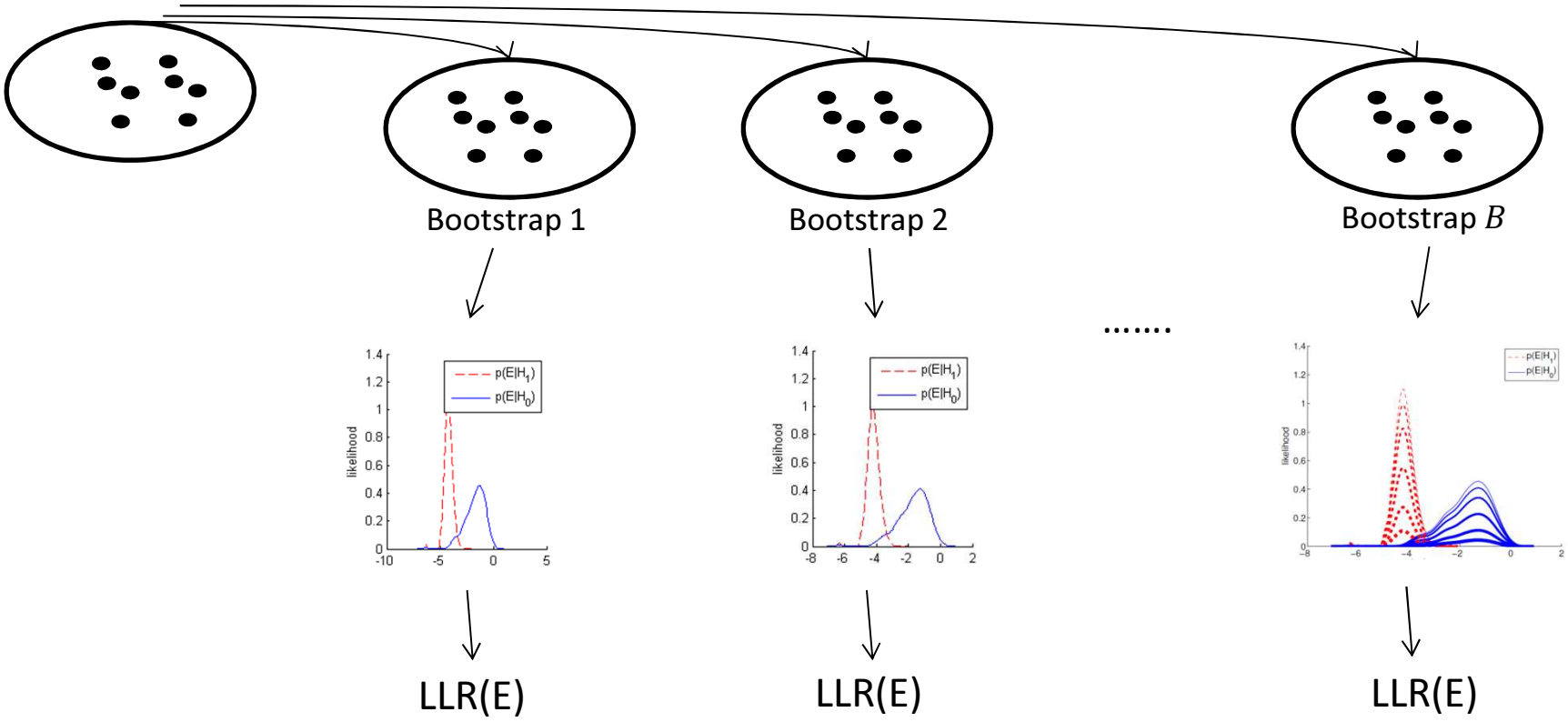

likelihood
score


5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
likelihood
−6
−4
−2
0
2
4
6
scores

Step 1: Bootstraps



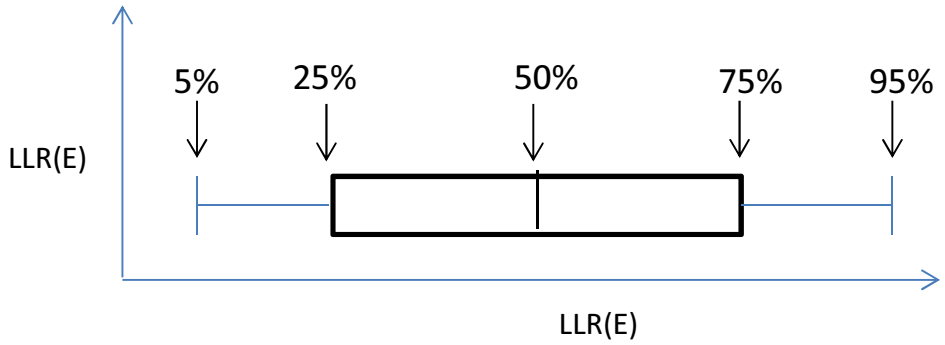
Step 2: H_0 and H_1 distributions



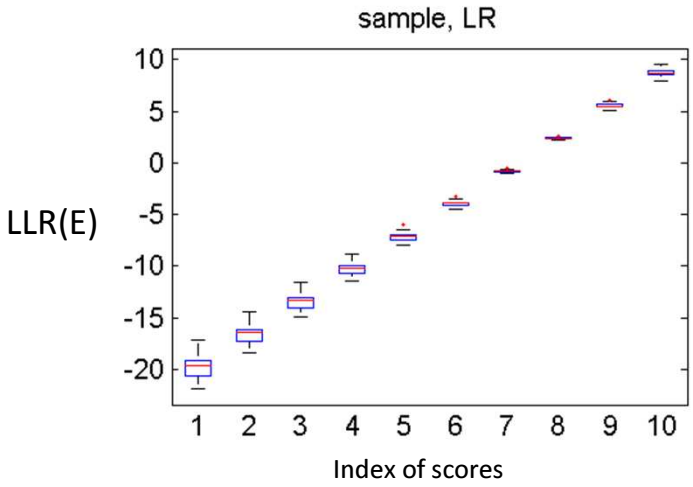
Step 3: LLR(E)

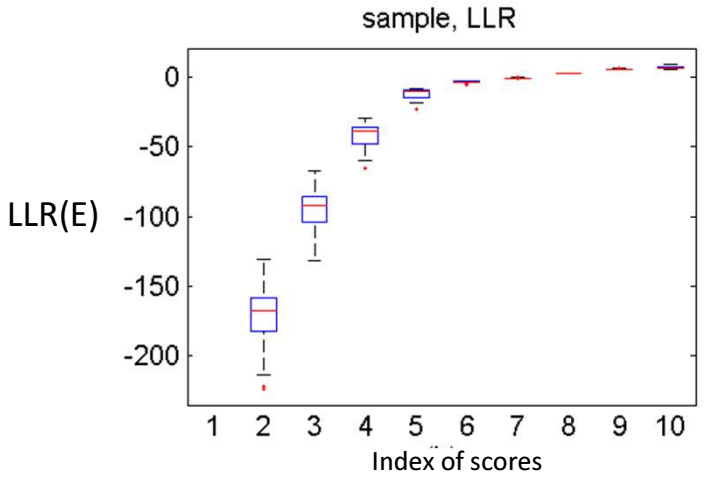
Logistic Regression	Density Estimation
$LR(E) = f_{lr}(E\|\theta_b)$	$LLR(E) = \log \frac{f_{llr}(E\|\theta_b^0)}{f_{llr}(E\|\theta_b^1)}$

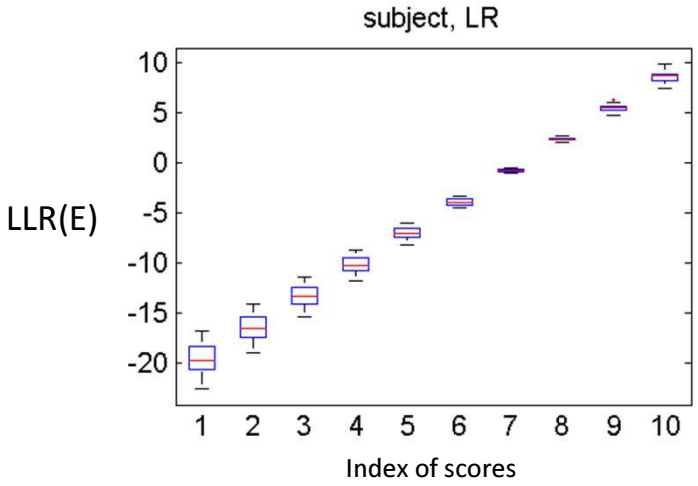
Step 4: Results

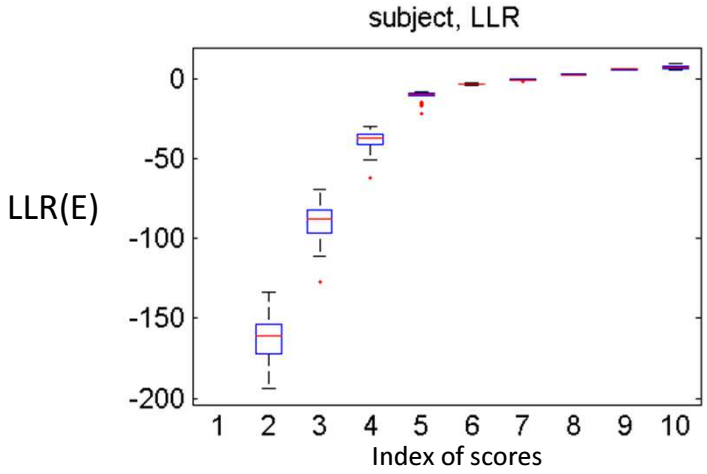


Experimental Results

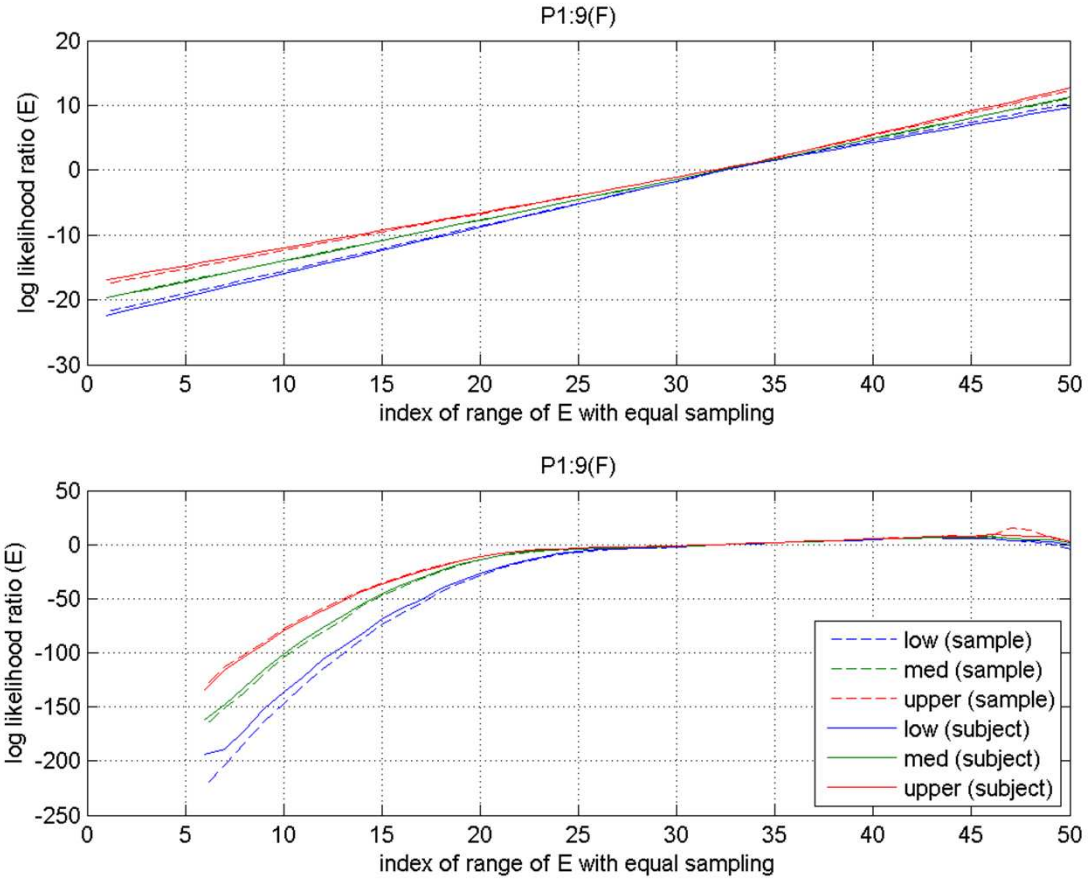








Confidence Intervals



Summary and Future Work

- Logistic Regression (LR) is better than Log-likelihood Ratio (LLR)
  - because the range of evidential strength is more compatible with what we expect.
- Bootstraping at subject level gives a larger confidence than bootstrapping at sample level → consistent with findings in [10].
- Future work: Apply the LR calibration to the proposed Antiforensic-resistant Framework as we introduced in [20].

20. N. Poh, N. Suki, A. Iorliam, and A. T. S. Ho, "Antiforensic likelihood ratio computation: A case study using fingerprint biometric," 2014.

Thank you ☺