The Confidence Interval for the Likelihood Ratio with Application to Biometrics

Larry Tang

George Mason University & National Institute of Standards and Technology

Joint work with *Elham Tabassi (NIST)* and *Xiaochen Zhu (GMU)*

Outline

- Evidence interpretation
- Error rates, receiver operating characteristic curve, and likelihood ratio
- Confidence interval of likelihood ratio on NIST datasets
- Conclusion and future work

Evidence Interpretation

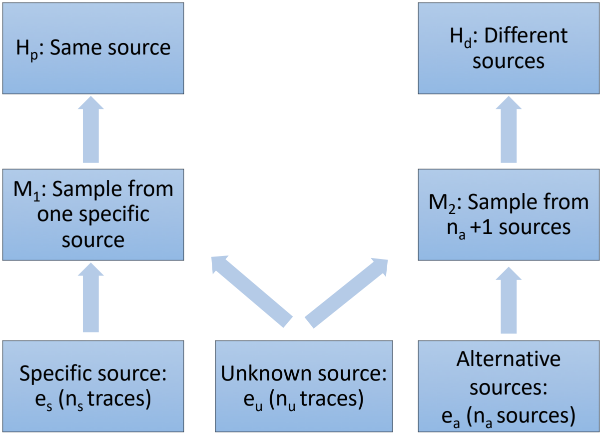
- The forensic source identification – inferential analysis to identify the origin of a collection of forensic evidence
- Summarization of the observed evidence relative to the prosecution and defense propositions
- Forensic scientists: interested in source level propositions, sometimes activity level propositions
- Court system: offense level propositions concerning the guilt or innocence of the defendant

Sources

Three subsets of objects:

- e_s: Set of objects associated with a specified source (person, window, ...)
- e_u: A set of trace objects from an unknown source
- e_a: Collection of sets of objects from alternative sources

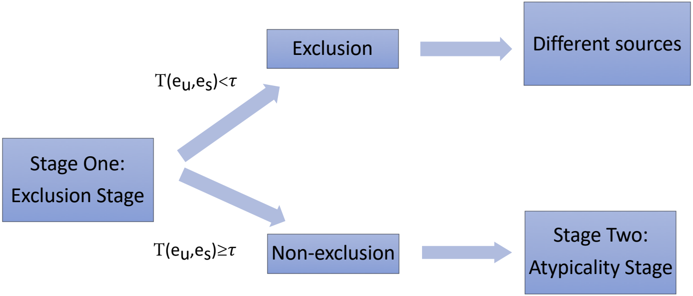
Propositions and Sources



Approaches for Summarizing the Evidence

- Bayesian methods (Lindley, 1977) – trace evidence
- Two-stage approach (Parker, 1966) – trace evidence
- Score-based methods (Royall,1997) – pattern and trace evidence

Two-stage Approach



Likelihood Ratio

- Error rates
  - Exclusion: random non-match probability (RNMP), or chance of incorrect non-match
  - Non-exclusion (or inclusion): random match probability (RMP), or chance of incorrect match
- Likelihood ratio (LR) based on binary decision of $T(e_s, e_u) >$ a threshold, τ:

$$LR = \frac{Pr(T(e_s, e_u) > \tau | H_p)}{Pr(T(e_s, e_u) > \tau | H_d)} = \frac{1 - RNMP(\tau)}{RMP(\tau)}$$

- Similar to diagnostic LR: accuracy of a diagnostic test which has positive and negative results
- Well-studied positive LR in diagnostic medicine: sensitivity/(1-specificity)

Forensic Error Rates

- Similarity scores for the i^{th} within-source comparison: $T_{s,i}, i = 1, \ldots, m$ follows F_{θ_s}
- Similarity scores for the j^{th} between-source comparison: $T_{d,j}, j = 1, \ldots, n$, follows F_{θ_d}
- Random non-match probability:

$$RNMP(\tau) = P(T_{s,i} \leq \tau) = F_{\theta_s}(\tau)$$

- Random match probability:

$$RMP(\tau) = P(T_{a,j} > \tau) = 1 - F_{\theta_d}(\tau)$$

ROC Curves for Forensic Error Rates

- ROC curve plots (1-RNMP) versus the RMP as the threshold point τ for determining a "match" varies from $-\infty$ to $+\infty$.
- Let t be $RMP(\tau)$, and $R(t)$ is $1 - RNMP(1-t)$
- ROC curve $R(t)$:

$$R(t) = 1 - F_{\theta_s}(F_{\theta_d}^{-1}(1-t))$$

- The derivative of the ROC curve closely related to likelihood ratio: the instantaneous change in the 1-RNMP in a unit change of RMP

Relationship between ROC and LR

An illustration of the relationship between ROC and LR:

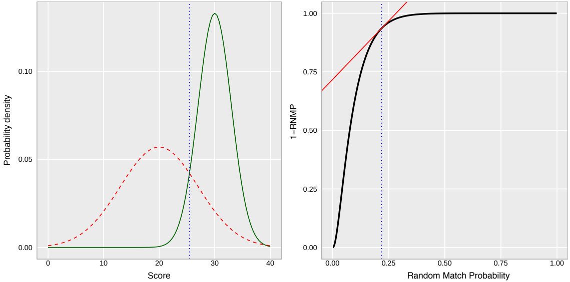


Figure: Left panel: dash curve – different-source scores, solid curve – same-source scores; right panel: solid black curve – ROC curve

Smooth ROC curve – Parametric Method

- Assume after Box-Cox power transformation, $F_{\theta_s} \sim N(\mu_s, \sigma_s^2)$ and $F_{\theta_d} \sim N(\mu_d, \sigma_d^2)$
- RNMP and RMP:

$$RNMP(\tau) = \Phi(\frac{\mu_s - \tau}{\sigma_s}), \quad 1 - RMP(\tau) = \Phi(\frac{\mu_d - \tau}{\sigma_d})$$

- The resulting binormal ROC curve :

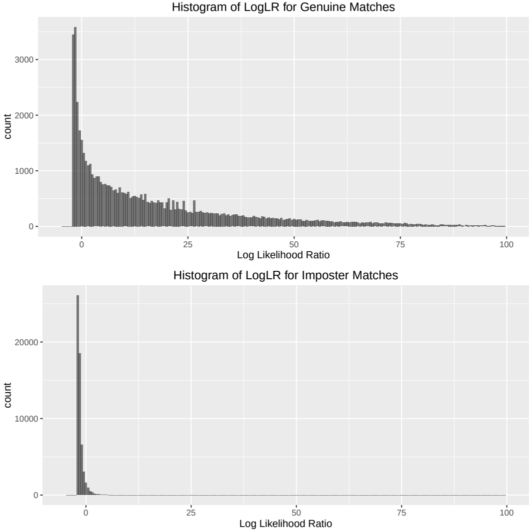
$$R(t) = \Phi(\frac{\mu_s - \mu_d}{\sigma_s} + \frac{\sigma_s}{\sigma_p}\Phi^{-1}(t))$$

- Explicit expression for LR estimate and its confidence interval

NIST SD4 Data

- NIST Special Database 4 (SD4)
- SD4 database contains 512-by-512-pixels gray scale fingerprint images
- Two representations for each finger – rolled impressions of the finger
- Bozorth matcher was run on all pairs of fingerprints from SD4 database

Histograms of NIST SD4 Data



Confidence Intervals of Log(LR) for SD4

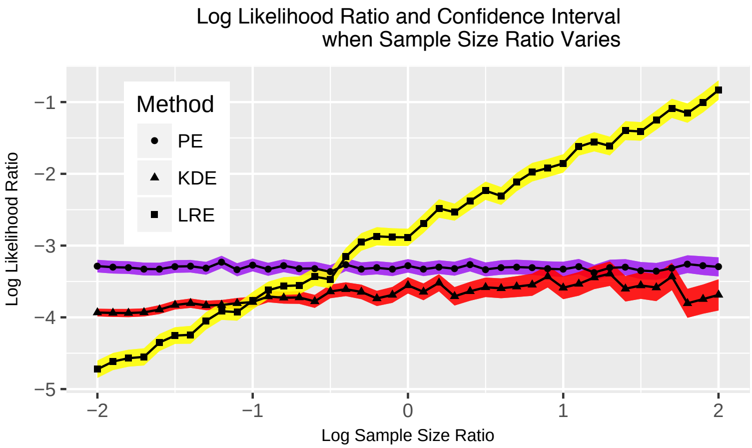
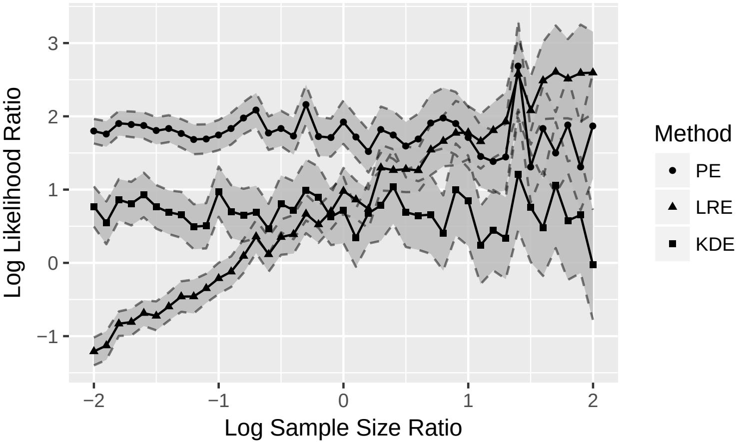


Figure: PE – parametric, KDE – kernel density estimation; LRE – logistic regression estimation (Zhu, Tang, Tabassi, 2017 IJCB)

Likelihood Ratios for Facial Recognition

- The Good, the Bad, and the Ugly Face Challenge Problem (Phillips, et al, 2012)
- Frontal face images taken with a digital single-lense reflex camera
- The data set has three categories, which are "good", "bad", and "ugly", based on the quality of the images
- The comparison scores measures characteristic difference

Confidence Intervals of Log(LR) for Facial Recognition



Conclusion and Future Work

- Sampling variability of likelihood ratio for fingerprint and facial recognition data
- Paradigm for the reasoning about the source of traces based on error rates
- Characterize the uncertainty about estimated forensic error rates

Acknowledgement

- NIST forensic program
- National Institute of Justice (Collaborators: Danica Ommen, Chris Saunders, Elham Tabassi, and Don Gantz)

Acknowledgement

- NIST forensic program
- National Institute of Justice (Collaborators: Danica Ommen, Chris Saunders, Elham Tabassi, and Don Gantz)

Thank you!