csafe

Center for Statistics and
Applications in Forensic Evidence

www.forensicstats.org

Statistical Analysis of User-Event Data for Digital Forensics

Chris Galbraith
Advised by Padhraic Smyth

University of California, Irvine

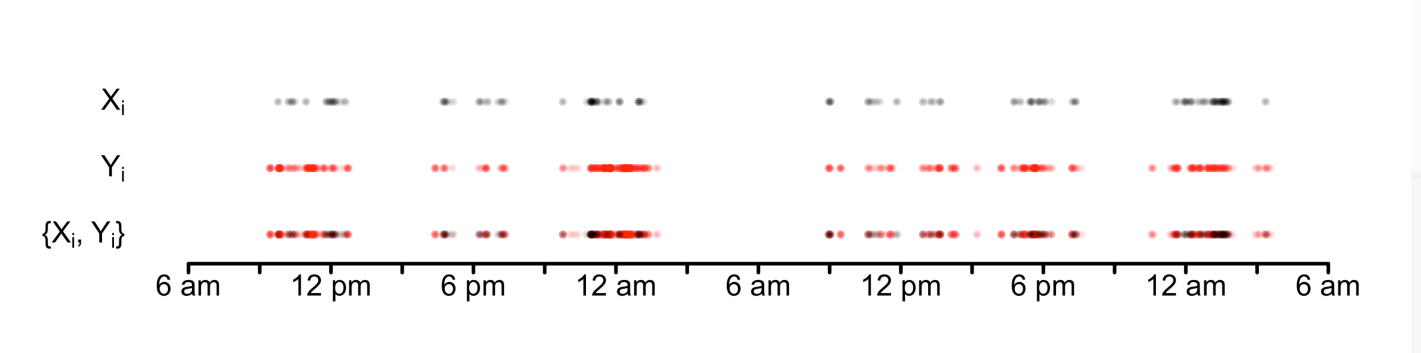
csafe
Center for Statistics and Applications in Forensic Evidence

Project Goals

1. Develop statistical methodologies to address questions such as
   - Are two event streams from the same individual or not?
     - Galbraith, C., & Smyth, P. (2017). “Analyzing user-event data using score-based likelihood ratios with marked point processes.” *Digital Forensic Research Workshop (DFRWS),* in press.
   - Are there unusual and significant changes in behavior?
2. Develop testbed data sets to evaluate these methodologies
3. Develop open-source software for use in the forensics community

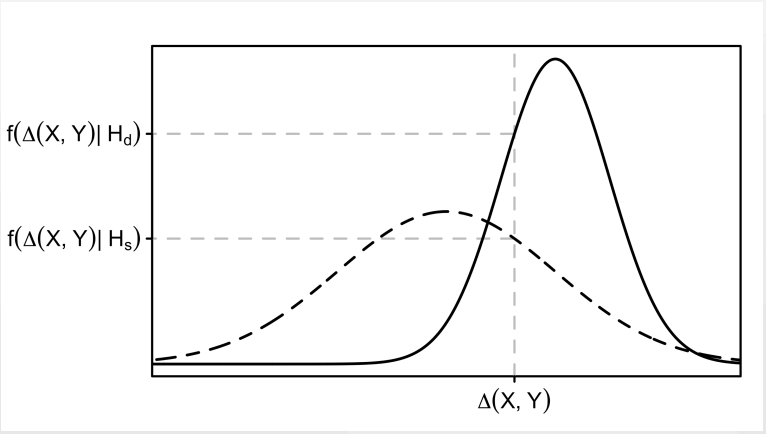
csafe
Center for Statistics and Applications in Forensic Evidence

- Use indices from *marked point process* literature as score functions that measure similarity between event streams



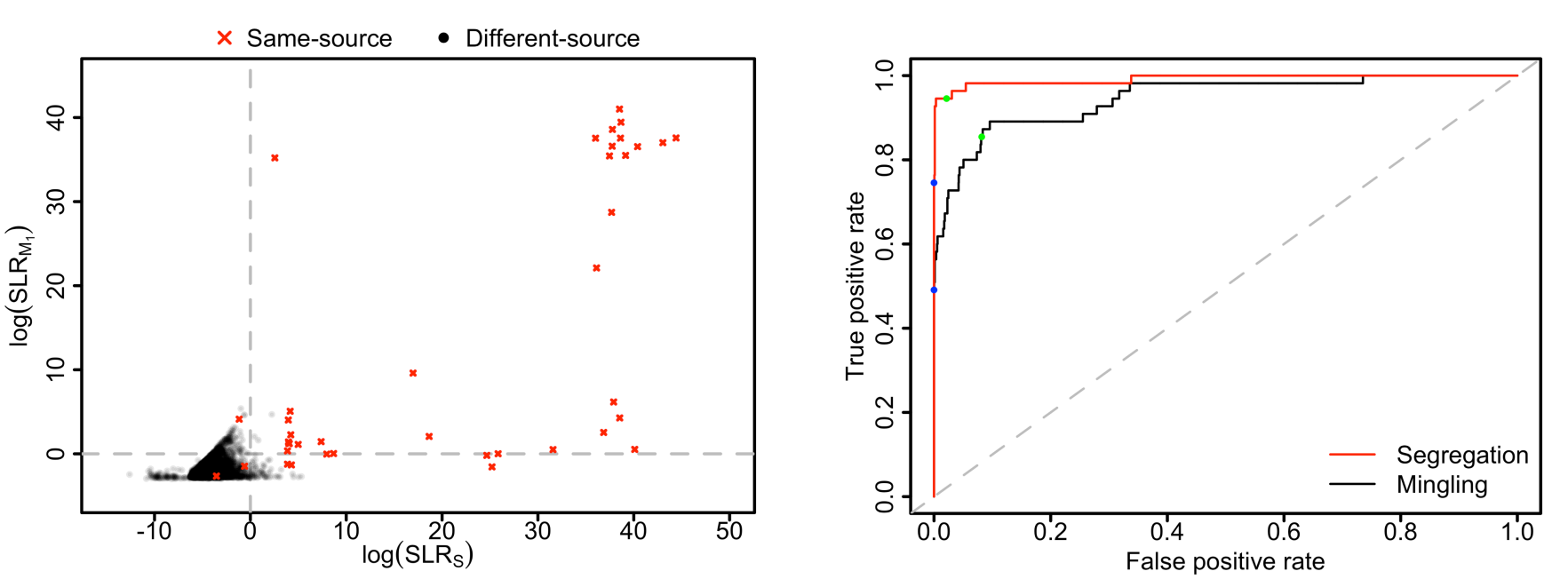
- Construct *score-based likelihood ratios* using these indices:

$$SLR_{\Delta} = \frac{\hat{f}(\Delta(X,Y)|H_s)}{\hat{f}(\Delta(X,Y)|H_d)}$$



csafe
Center for Statistics and Applications in Forensic Evidence

- Data: event streams (Facebook & non-FB web browsing) from 55 students measured for one week
- SLR_{Δ} estimated for each pairwise combination of individuals and their event streams, based on data from all *other* individuals



csafe
Center for Statistics and Applications in Forensic Evidence

- **Conclusions:**
  - Marked point processes can provide a general statistical framework for the analysis of human-generated event data in digital forensics
  - SLRs based on mpp indices showed significant promise in addressing whether two event streams came from the same or different individuals, using a real-world data set of web browsing events
  - Results are *only for one specific data set* and may not generalize to others
- **Future Work:**
  - Broaden the scope of the statistical methodology
    - e.g., inter-event times & multiple types (>2) of event streams
  - Gather additional real-world testbed data sets
  - Disseminate ideas at technical forensics conferences (DFRWS & ICFIS)
  - Investigate robustness and applicability of randomization methods
  - Investigate theoretical limits of detectability & different types of dependence

csafe
Center for Statistics and Applications in Forensic Evidence