Event histories recording user activities are routinely logged on devices such as computers and mobile phones. For a particular user these logs typically consist of a list of events where each event consists of a timestamp and some metadata associated with the event. As digital devices become more prevalent, these types of user event histories are encountered with increasing regularity during forensic investigations. As an example, an investigator might be trying to determine if two event histories, corresponding to different usernames, were in fact generated by the same individual.

**SCORE-BASED LIKELIHOOD RATIO**

Let $E = \{X, Y\}$ where $X$ is a set of observations for a reference sample from a known source, and $Y$ is a set of observations of the same features as $X$ for a sample from an unidentified source [1]. The likelihood ratio is the ratio of the probability of observing the evidence $E$ under two competing hypotheses (that the samples come from the same source, $H_s$, or different sources, $H_d$). The LR arises in the application of Bayes’ theorem to this situation:

$$\frac{Pr(H_s|E)}{Pr(H_d|E)}$$

Instead of modeling the probability of observing the evidence $E$ directly, one can model the probability density of a score function $\Delta$ that measures the similarity between the samples $X$ and $Y$. This yields the score-based likelihood ratio:

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

**BIVARIATE POINT PROCESSES**

The event stream for the $i^{th}$ user can be viewed as one-dimensional bivariate point processes $M_i = \{X_i, Y_i\}$ where $X_i = \{t_j : m(t_j) = 0 \text{ for } j = 1, \ldots, n_i\}$ is the sub-process of events of type $x$ (similar definition for $Y_i$). We use indices (coefficient of segregation, S, and mingling index, M), as score functions.

$$S(X_i, Y_i) = 1 - \frac{p_{xy} + p_{yx}}{p_x p_y}, \quad M_k(X_i, Y_i) = \frac{1}{k} \sum_{j=1}^{n_i} \sum_{t_i \in M} I[m(t_i) \neq m(z_i(t_i))] \in [0,1]$$

**RESULTS**

Score-based likelihood ratios based on marked point process indices have potential to perform well in quantifying strength of evidence for user-event data. In particular, segregation and mingling were discriminative score functions for web browsing event streams. Note that results obtained only for specific data set and may not generalize to others.

**FUTURE WORK**

- Randomization methods (given only one set of event streams)
- Inter-event times
- Multiple types (>2) of events
- Change detection in this context

**REFERENCES**