

# **Master thesis project at the Netherlands Forensic Institute:**

## **Development and validation of Likelihood Ratio methods**

In the forensic scientific community, it is widely accepted to present a likelihood ratio (LR) as a measure for the strength of evidence in court [1]. This quantity is also reported in contributions from the Netherlands Forensic Institute to the Dutch legal system.

A LR is defined as the ratio of two probabilities. The numerator is the probability of obtaining the evidence when the prosecutor's scenario is true. The denominator is the probability of obtaining the evidence when the defense scenario is true. A LR of 1 represents neutral evidence, while a LR larger than 1 supports the prosecutor's hypothesis and a LR between 0 and 1 supports the defense hypothesis. Currently, for the vast majority of evidence modalities, conclusions on the value of evidence in reports to be used in court use a verbal expression for the LR. The main reason for this is that it is still a scientific challenge to obtain a numerical value for the LR.

There are a number of evidence modalities where scientific progress is made in this respect, for example the evidence modalities finger mark comparison, speech comparison, facial comparison and DNA comparison.

Design of LR-methods starts with a training sample dataset that contains the variability of the evidence given the prosecutors' and the defense scenario. It is common practice to model the population distribution of the evidence given a hypothesis using this training dataset as a sample. To measure the success of this modeling approach, a validation study has to be undertaken. Current standards on validation of LR methods recommend using an independent validation dataset (which is also a sample from the population under the prosecutors and the defense hypothesis) [2]. However, often, obtaining a new validation dataset is an expensive, time-consuming process. Therefore, for example in the field of finger mark comparison, a leave-one-out cross validation strategy is pursued [3]. The question is whether using this strategy is expected to lead to different results than using an independent validation dataset. A comparison study on the effect of using a leave-one-out cross validation strategy versus an independent validation dataset is the topic of this thesis.

### ***Study strategy***

A crucial aspect of the design of a LR method is calibration [4]. When LRs from a LR-method are not well calibrated, they represent the wrong number, which could lead to disproportionate rates of misleading evidence. Calibration proceeds via two approaches, with the common goal to obtain a LR that is in accordance with data. One approach is to use the training sample data to find the transformation function of evidence (expressed as a univariate, continuous parameter, for example a comparison score) to LR directly, the other is to estimate the two probability distributions of the evidence given a scenario from the training sample. In this case the LR for certain evidence is obtained by dividing the modeled probability densities at the values corresponding to the evidence.

An important goal of a validation study is to measure the calibration, and this study is focused on this quantity. For the measure of calibration a test dataset of LRs (obtained from the validation dataset or a leave-one-out strategy) is obtained. In the forensic scientific literature, calibration is measured by ECE [4] and (soon) experimental LR calibration plots [5], which will both be used to measure the calibration.

The strategy for this study is as follows. The NFI provides sample data from the population of evidence for typical prosecution and defense scenarios for several evidence modalities (e.g. speech and glass fragments evidence). By a random process, these datasets will be split. The first half is used for calibration. The process for measuring calibration is manipulated in the study. In the first condition, the second half of the dataset is used to obtain ECE and empirical LR calibration plots. In the second condition, the second half of the data is not used, but instead a leave-one-out cross validation scheme

is used to obtain ECE and empirical calibration plots. Data describing the variability in the results is obtained by resampling methods, each time obtaining a new random split from the original data set.

## ***Outcome of the study***

The outcome of this study is a comparison of both validation strategies, in terms of the variability in ECE and LR empirical calibration plots. It is the intention to publish the results of the study in an international forensic scientific journal.

## ***Background of the student***

The student can have a background in either statistics; electrical engineering or computer science, but with an affinity for statistics and have some knowledge of statistical programming (R or MATLAB). Furthermore he or she must have an interest in forensic science.

## ***Interested?***

If you are interested, please contact:

- Prof. Dr. Didier Meuwly ([d.meuwly@utwente.nl](mailto:d.meuwly@utwente.nl)) or Prof. Dr. Raymond Veldhuis ([r.n.j.veldhuis@utwente.nl](mailto:r.n.j.veldhuis@utwente.nl)) from the SCS group, or
- Dr. Arnoud den Boer ([a.v.denboer@utwente.nl](mailto:a.v.denboer@utwente.nl)) from the SOR group.

The supervision at NFI will be by Dr. Peter Vergeer ([p.vergeer@nfi.minvenj.nl](mailto:p.vergeer@nfi.minvenj.nl)).

## ***References***

- [1] I. W. Evett, Towards a uniform framework for reporting opinions in forensic science casework, *Science & Justice* 38, (1998) 198-202.
- [2] G. Zadora, A. Martyna, D. Ramos, and C. G. G. Aitken, Performance of likelihood ratio methods, Ch. 6 in *Statistical analysis in forensic science: evidential value of multivariate physicochemical data*, first ed., (John Wiley and Sons, Chichester, UK, 2014).
- [3] A. J. Leegwater, D. Meuwly, M.S. Sjerps, P. Vergeer, I. Alberink, Validation of a score-based likelihood ratio system for fingerprint comparison in forensic casework, (2014), in prep.
- [4] D. Ramos and J. Gonzales-Rodriguez, Reliable support: Measuring calibration of likelihood ratios, *Forensic Science International* 230, (2013) 156-169.
- [5] P. Vergeer, A. Bolck, I. Alberink, M.S. Sjerps, R.D. Stoel, D.A. van Leewen, Measuring calibration as a function of likelihood ratio: using the PAV transform as a visual aid, (2014), in prep.