Current challenges in statistical DNA evidence evaluation

Cereda, G.

Part III

Discussion

Chapter 9

Discussion and conclusion

This thesis is the result of a five-year study conducted in between the Faculty of Criminal Justice of Lausanne and the Mathematical Institute of Leiden. This hybrid background gave rise to a research that contributed to improving both the domains.

9.1 Contribution to the practice of Forensic Science

The original aim of the thesis was to develop a Bayesian evaluative framework for the results obtained with DIP-STR technology, which in turns constitutes an answer to the problem of extremely unbalanced mixtures. Based on the use of graphical models, it allows using the results obtained with the DIP-STR technology in a legal context, inasmuch it leads to the calculation of the likelihood ratio for any observation.

Additionally, the thesis provides several solutions to deal with the rare type match problem. Generalizations of the Good-Turing estimator and of the discrete Laplace models are used in a frequentist context. A method based on the use of a Bayesian nonparametric prior, and a revisiting of the classical Dirichlet-multinomial model, are proposed and discussed in a Bayesian framework. At the best of our knowledge, this thesis constitutes the first introduction of Bayesian nonparametric prior to forensic applications. Given the satisfying results, we are confident this is only the beginning.

The lemma described in Section 2.5 and proved in Chapter 7 is of very broad application and very useful in many forensic cases. In fact, it allows one to simplify the calculation of the likelihood ratio in all the situations in which prosecution and defence agree on the distribution of part of the available data (for instance, they both see the reference database as a random sample from the population), but they disagree on the distribution of the rest of the data (for instance, the correspondence between the DNA profile of the suspect and of the crime stain is a random event according to the defence, while it is a sure event according to the prosecution).

9.2 Contribution to the Philosophical point of view

Bayesianism dates back of at least one century with, among others, De Finetti (1931, 1937), Wald (1949), and Savage (1954). It is commonly perceived as an opponent to classical frequentist statistics. The use of Bayesian statistics is, according to many, the natural choice for a statistician working in the legal framework, as testified by important pieces of literature such as Lindley (1977b,a, 1978), Robertson and Vignaux (1995), and Aitken and Taroni (2004).

The use of frequentist methods to assess the likelihood ratio may be seen as less coherent, since the likelihood ratio is then used within the Bayes' theorem context, as the way to update prior odds to posterior odds. However, many frequentists statisticians are interested as well in the likelihood ratio, seen as a tool to measure the evidential value of data, independently of the Bayes' theorem.

In order to study the rare type match problem, we came in contact with both the Bayesian and the frequentist approaches to likelihood ratio assessment. We believe that often literature proposes hybrid solutions, passed off as Bayesian, such as the plug-in Bayesian approximations, which in fact can be seen as a compound of the two approaches. Hence, we felt the need to set up a formal distinction between the two. In particular, we wanted to emphasise that the frequentist approach can be seen as a Bayesian approach with special prior over the nuisance parameters. The difference among the two lies in the definition of the probability.

Lastly, the Bayesian plug-in method proposed in much forensic literature, which consists of estimating the unknown allelic frequencies (nuisance parameter of the model) using the mean of their posterior distribution after the observation of a database, is discussed and compared with the full Bayesian approach that integrates out the nuisance parameters. The latter is often not more difficult, but most of the time the likelihood ratios using the two methods do not differ substantially. The Lemma proposed allows one to calculate the full Bayesian likelihood ratios by calculating the posterior expectation of a simple function of the parameter, instead of performing classical marginalization steps.

9.3 Future perspective

The interpretative framework developed for DIP-STR results from a mixed trace, took as assumption that the number of contributors is known, and equal to two. However, DIP-STR markers can also be used to investigate situations in which a fixed number of contributors cannot be agreed with certainty, extending the modular assessment procedures. The uncertainty about the number of contributors is to be taken into account as a further variable when evaluating DIP-STR profiling results. The question of how many individuals have contributed to a given mixture is a general issue that is independent of the type of analysis which is chosen (i.e., traditional STR or DIP-STR), but the fact that DIP-STR alleles have a two-dimensional set of labels for the alleles (two different possible DIP alleles for each STR alleles), could potentially provide more discriminative power over the number of contributors. This can be explored extending the model derived in the first part of the research (the

OOBN constructed to model DIP-STR results), taking stands on the structure proposed in Biedermann et al. (2011b). Through simulations, it is conceivable to compare this method against the classical methods (typically STR and Y-STR) to see if there is an advantage in using DIP-STR markers when the number of contributors in not known in advance.

Despite the numerous results obtained we felt that much more can be done for the rare type match problem. Especially regarding Bayesian nonparametric methodologies, we are just at the dawn: new kinds of nonparametric priors can be studied and used instead of the two-parameter Poisson-Dirichlet distribution, new methods to choose their hyperparameters can be adopted. Of particular interest for the rare type match problem, is the 'k-method' of Brenner (2010). We believe that a rigorous definition of the statistical background is missing, and one of the next steps for this research could be that of better formalising this solution.

9.4 Conclusion

This thesis represented an opportunity to study many crucial problems concerning forensic statistics. This five-year study uncovered many interconnections between them, but there is no clear-cut answer, applicable to all problems. Each one needs a tailored solution, and many aspects should be taken into account and weighed. Issues such as data reduction, uncertainty assessment, hybrid approximations, are rarely discussed and studied in literature. Moreover, there is a big contraposition of Bayesian and frequentist approaches to probability definition. Both approaches have advantages and disadvantages and both need further refinement and improvement. However, the Bayesian approach is more appropriated for legal and forensic reasoning (and easier to explain to lay-persons), even though the choice of the priors is a delicate and crucial problem, often underestimated.